Bayesian Correlated Factor Analysis for Spatial Data
Maura Mezzetti
Dipartimento di Studi Economico-Finanziari e Metodi Quantitativi,
Università Tor Vergata, Roma
e-mail: maura.mezzetti@uniroma2.it

Abstract: A hierarchical Bayesian factor model for multivariate spatially correlated data is proposed. The idea behind the proposed method is to search factor scores incorporating a dependence due to a geographical structure. The great flexibility of the Bayesian approach bears directly on the problem of parameter identification in factor analysis and furthermore on the inclusion of our prior opinion about adjacent regions having high correlated observable and latent variables. The underlying idea taken from Rowe (1998) is the introduction of separable covariance matrix for the observation vector $X' = (x_1, x_2, \ldots, x_N)$, ($N \times p$ vector where $N$ is the number of observations and $p$ number of variables) so that $\text{var}(X) = \Phi \otimes \Psi$ where $\otimes$ denotes the Kronecker product. The advantage of introducing a separable covariance matrix is that we can now interpret the matrix $\Phi$ as the between-observations covariance matrix, and the matrix $\Psi$ as the within-observations covariance matrix. In the paper, different prior distributions for $\Phi$, the between-observations covariance matrix, are investigated and both informative and uninformative priors are explored. High correlation between regions can be due to geographic distance or cultural affinity. The model extends a methodology for temporal dependence pattern proposed by Mezzetti and Billari (2005). A Gibbs sampling algorithm is implemented to sample from the posterior distributions. The methodology will be illustrated through the analysis of epidemiological data.

Keywords: Bayesian inference, factor analysis, correlated factor loadings, spatial data.

1. Introduction

Factor analysis is a powerful statistical tool for describing and modeling the underlying structure in multivariate data. The analysis is widely used in situations where observed variables can be explained by a smaller number of unobservable factors and where underlying theoretical quantities cannot be measured directly.

Most proposals in the factor analysis literature assume that the data represent random, independent samples from a multivariate distribution. This is not necessarily a good assumption for all types of multivariate data. For certain types of data, observations appear in a specific order, and it is no longer permissible to exchange the order of observations without a fundamental change in the outcome (Basilevsky, 1994). Time series data and panel data, for example, are in conflict with the previous assumption. Time series factor analysis and dynamic factor analysis were introduced in this direction.

Furthermore, when several variables are measured at one specific location over a spatial area, they are often correlated across the locations due to geographic similarities of the different locations. This type of multivariate spatially referenced data is commonly seen in public health or environmental research. Multivariate spatial data are another example of data in conflict with the assumption of independence between observations.
The need to analyze or summarize multivariate data that are spatially distributed suggests the utility of a hierarchical model that can incorporate spatial covariation.

As it will be illustrated in the following section, in the usual common factor model identifiability problems are encountered. The Bayesian approach bears directly on the problem of parameter identification, by incorporating proper prior information. The Bayesian factor analysis model incorporates available knowledge regarding the model parameters in the form of prior distributions obtained either subjectively from substantive experts or from previous experiments. Moreover, the Bayesian approach to factor analysis removes the ambiguity in the choice of rotation procedures.

The proposed method is a Bayesian approach to factor model dealing with correlation between observations. The innovation comes from a work by Rowe (1998) and consists on the introduction of separable covariance matrix for the observation vector $X' = (x_1, x_2, \ldots, x_N)$, ($N \times p$ vector where $N$ is the number of observations and $p$ number of variables) so that $\text{var}(X) = \Phi \otimes \Psi$ where $\otimes$ denotes the Kronecker product. The introduction of a separable covariance matrix represents the main innovation and its great advantage consists on the interpretation: $\Phi$ in now the between-observations covariance matrix, and the matrix $\Psi$ is the within-observations covariance matrix. The idea is to investigate different structure for the prior distribution for the matrix $\Phi$, between-observations covariance matrix, exploring both informative and uninformative priors, incorporating our prior opinion about correlation between adjacent regions. Mezzetti and Billari (2005) propose a Bayesian model for analysis of geographic panel data able to handle the dependence between the observations extending Rowe’s idea to multiple parallel time series and to estimate factor scores that account for a temporal dependence.

Interest in the analysis of geographical variation in rates of disease (or mortality) is increasing since it is fundamental in the formulation and validation of aetiological hypotheses. Recently, the availability of local geographically indexed health and population data, together with advances in computing and geographic information systems, has encouraged the analysis of health data on a small geographic scale (Elliott et al., 2000). An application to cancer disease in Scotland will explore the potentiality of the proposed methodology. Latent factors will be extracted from a multivariate series of data consisting, on one hand, on disease distributions of different cancer sites and, on the other hand, on covariates representing possible risk factors for cancer. Through a Bayesian approach is possible to incorporate our prior opinion about spatial pattern in disease risk, i.e. the tendency for geographically close areas to have similar disease rates. Through a Bayesian approach, we can look for latent factors explaining different distribution of the disease by defining a prior distribution for matrix $\Phi$ that attributes a high correlation to two regions, (or census tracts), that are close to each other.

The idea behind the method proposed is to extend Rowe’s proposal to spatial correlated data. A detailed description of method proposed by Rowe (2003) is contained in Section 2, and in Section 3 our proposal is shown. After an illustration of the computational aspects in Section 4, application to cancer epidemiologic data will be shown in Section 5. Section 6 contains final conclusions and remarks.
2. Correlated Bayesian Factor Model

We begin this section by recalling classical and Bayesian factor analysis, in order to introduce the model developed in Rowe (2003). This will simplify the illustration of our approach in the following section.

Factor analysis is used either for data reduction or to explain the observed relationship among a set of observed variables in terms of a smaller number of unobserved variables or latent factors which underlie the observations. This structure will aid in the interpretation and explanation of the process that has generated the observations.

Let \( x_i \) denote the \( p \)-vector of observation on subject \( i \) of \( p \) random variables. A factor analysis is generally based on the following model:

\[
(x_i | \mu, \Lambda, f_i, m) = \begin{pmatrix} \mu \\ \Lambda \\ f_i \\ \varepsilon_i \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (1)
\]

\( \mu \) is a \( p \)-dimensional unobserved population mean vector, 
\( \Lambda \) the \( p \times m \) matrix of unobserved constants called the factor loadings matrix, 
\( f_i \) a \( m \)-dimensional vector of unobservable “common” factor scores for the \( i \)-th subject, and 
\( \varepsilon_i \) a \( p \)-dimensional vector of “specific” factors or disturbance terms of \( i \)-th subject on \( p \) variables.

In the traditional factor analysis model, the errors are assumed to be normally distributed with mean 0 and (in the non-Bayesian model) diagonal covariance matrix \( \Psi \). The main reason in the assumption of uncorrelatedness of all component of \( \varepsilon \) is that any correlation observed among the components of \( X \) is accounted for by the common factors. In the Bayesian model, by permitting component \( \varepsilon \) to be correlated we account for the possibility of specification error; that is, omission of one or more factors from the model.

The parameters \((\mu, \Lambda, f, \Psi)\) in the model are unknown and thus require estimation. The number of factors \( m \) can be determined based on underlying theory and previous studies. Different rules exist for the choice of number of factors in the non-Bayesian literature (such as a scree test or percent variation), while a probabilistic approach is used in Bayesian context. The estimate of the population mean \( \mu \) is easily found by maximum likelihood and coincides with sample mean, see (Lawley, 1940). From now on, to simplify calculations but without lack of generality, we will assume that \( x \) vector has zero mean, moreover if \( x \) is centered and scaled, than \( \Lambda \) is a correlation matrix between the \( x \) and \( f \).

The factor scores can be considered either random vectors or nonrandom vectors. The model is overparameterized, the likelihood does not have a maximum, and we cannot reach maximum likelihood estimates through differentiating the log likelihood function. By assuming the factor scores not fixed, but random normally distributed variables with mean 0, standard deviation 1 and correlation \( R \), independent from error random variables \( \varepsilon_i \), it is possible to overcome unidentifiability problems. The variance and covariance matrix of observed vectors can be written as \( Var(x_i | \Lambda, \Psi, m) = \Lambda \Lambda' + \Psi \) and estimated by sample variance \( \hat{\Sigma} \). After adding constraints on the parameters (as for example \( R \) being the identity and \( \Lambda \) being columnwise orthogonal), we can get unique solutions.

The Bayesian approach bears directly on the problem of parameter identification, by incorporating proper prior information. If subjective information is introduced through the class of priors suggested, the usual rotational identification problem will disappear so
that all the parameters of the factor loading matrix will be identified. We follow here the approach of Press and Shigemasu (1997) to date considered the best and most complete reference on Bayesian factor analysis. We start from the model in (1), to obtain the likelihood we assume:

$$
\varepsilon_i \sim N(0, \Psi), \quad i = 1, \ldots, N
$$

assuming $E(\Psi)$ is diagonal to represent traditional beliefs of the model containing common and specific factors. While Lawley (1940) hypothesizes that the matrix is strictly diagonal, here we hypothesize that it is a full positive definite diagonal matrix on average. The likelihood for the observations can be written as the following matrix normal distribution

$$
p(X|F, \Lambda, \Psi, m) \propto |\Psi|^{-N/2} \exp \left( -\frac{1}{2} \text{trace} \left( (X - FA')\Psi^{-1}(X - FA')' \right) \right) \tag{2}
$$

where the $i$-th row of $X$ and $F$ are, respectively, $x_i'$ and $f_i'$, cited in (1).

In Press and Shigemasu (1997), natural conjugate families of prior distributions for parameters are used. The factor loadings are assumed to depend on the disturbance covariance matrix. The disturbance covariance matrix is assumed to be independent of the factor scores. The factor scores are assumed to be independent of the factor loadings and the disturbance covariance matrix. More specifically, the joint prior distribution has the following form:

$$
p(F, \Lambda, \Psi|m) = p(\Lambda|\Psi, m)p(\Psi)p(F|m),
$$

moreover, we have the following set of prior distributions:

$$
p(\Lambda|\Psi, m) \propto |\Psi|^{-m/2} \exp \left( -\frac{1}{2} \text{trace} \left( \Psi^{-1}(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' \right) \right), \tag{3}
$$

$$
p(\Psi) \propto |\Psi|^{-\nu/2} \exp \left( -\frac{1}{2} \text{trace} \Psi^{-1}B \right) \quad \nu > 2p, \tag{4}
$$

$$
p(F|m) \propto \exp \left( -\frac{1}{2} \text{trace} F'F \right). \tag{5}
$$

with $\Psi > 0$, $H > 0$, and $B > 0$ and a diagonal matrix, such that $E(\Psi|B)$ is diagonal to represent traditional assumption of common and specific factors.

Straightforward posterior distributions are reached. More specifically, the factor scores given the factor loadings, the disturbance covariance matrix and the data are normally distributed as:

$$
p(F|\Lambda, \Psi, X, m) \propto \exp \left( -\frac{1}{2} \text{tr} \left( (F - \tilde{F})(I_m + \Lambda'\Psi^{-1}\Lambda)(F - \tilde{F})' \right) \right), \tag{6}
$$

where $\tilde{F} = X\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$. The conditional posterior of the factor loadings given the factor scores, the disturbance covariance matrix, and the data is normally distributed:

$$
p(\Lambda|F, \Psi, X, m) \propto \exp \left( -\frac{1}{2} \text{tr} \left( \Psi^{-1}(\Lambda - \tilde{\Lambda})(H + F'F)(\Lambda - \tilde{\Lambda})' \right) \right) \tag{7}
$$
where $\tilde{\Lambda} = (X'F + \Lambda_0H)(H + F'F)^{-1}$.

The conditional posterior density of the disturbance covariance matrix given the factor scores, the factor loadings, and the data is an inverted Wishart density:

$$p(\Psi|F, \Lambda, X, m) \propto |\Psi|^{-\frac{N+m+p}{2}}\exp\left(-\frac{1}{2}tr\left(\Psi^{-1}U\right)\right)$$

(8)

where $U = (X - F\Lambda)'(X - F\Lambda) + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

As in Bayesian inference, the expected values of the conditional posterior distributions (6), (7) and (8) are a weighted mean between prior expected values and maximum likelihood estimates. A Gibbs sampling is easily implemented in the three previous conditional posterior distributions (Rowe and Press, 1998).

In maximum likelihood factor analysis, the covariance matrix for the errors of the observations is supposed to be diagonal while in the Bayesian factor analysis it is assumed to be positive definite, but diagonal on average. In both models, the error vectors are assumed to be independent (or conditionally independent). We try to remove the previous assumption by first rewriting (1), as:

$$(x|\mu, \Lambda, f, m) = I_N \otimes \Lambda \quad \begin{bmatrix} f \\ (N \times N \otimes p \times m) \end{bmatrix} + \varepsilon$$

(9)

assuming the error has the following distribution:

$$\varepsilon \sim N(0, \Omega),$$

the innovative idea proposed by Rowe (2003) is to assume separable covariance matrix, that is $\Omega = \Phi \otimes \Psi$ ($\otimes$ indicate Kronecker product). If we let $\Phi$ be the identity matrix, we have the model in (1). In this way, $\text{var}(x_i|\Phi, \Psi, m, f, \Lambda) = \phi_{ii}\Psi$ and the covariance between rows $i$ and $j$ of $X$ is $\phi_{ij}\Psi$, while the covariance between columns $i$ and $j$ of $X$ is $\psi_{ij}\Phi$. This model will be explained more in details in next section where we will introduce our proposed model.

3. Model Proposed

Starting from model in (9), the likelihood for the observations can be written as the following matrix normal distribution

$$p(X|F, \Lambda, \Psi, \Phi, m) \propto |\Phi|^{-p/2}|\Psi|^{-N/2}\exp\left(-\frac{1}{2}trace\left(\Psi^{-1}(X - F\Lambda)'\Phi^{-1}(X - F\Lambda)\right)\right).$$

(10)

where again the $i$-th row of $X$ and $F$ are, respectively, $x_i$ and $f_i$.

We will use natural conjugate families of prior distributions for the parameters. The joint prior distribution is given by

$$p(\Phi, \Psi, F, \Lambda|m) = p(\Psi)p(\Phi)p(F|\Phi, m)p(\Lambda|\Psi, m).$$
where

\[ p(\Lambda|\Psi, m) \propto |\Psi|^{-m/2} \exp \left( -\frac{1}{2} \text{trace} \left( \Psi^{-1}(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' \right) \right), \quad (11) \]

\[ p(\Psi) \propto |\Psi|^{-\nu/2} \exp \left( -\frac{1}{2} \text{trace} \Psi^{-1}B \right) \quad \nu > 2p, \quad (12) \]

\[ p(F|\Phi, m) \propto |\Phi|^{-m/2} \exp \left( -\frac{1}{2} \text{trace} \Phi^{-1}FF' \right). \quad (13) \]

Again, \( \Psi > 0, H > 0, \) and \( B > 0 \) and a diagonal matrix, such that \( E(\Psi|B) \) is diagonal.

Basically, prior distributions for \( \Psi \) and \( \Lambda \) remain the same as in (4) and (3), indicating that dependence between observations does not affect prior opinions about factor loadings and between observation covariance matrix. Prior distribution on \( F \) changed from (5), since dependence between observation is reflecting on \( F \). Let \( f \) be a vector \( Nm \times 1 \), containing columns in \( F \), its covariance matrix is the Kronecker product of an identity matrix (expressing independence between factors) and \( \Phi \).

After definition of likelihood in (10), an appropriate prior distribution for the matrix \( \Phi \) needs to be defined; prior distribution will be searched such that leaves posterior distributions analytically tractable. We propose three different prior distributions for the matrix \( \Phi \). Letting \( p(\Phi) \) as unspecified for the moment, conditional posterior distributions are reached in a straightforward way. More specifically, the joint posterior distribution for the unknown parameters of interest is given by:

\[ p(F, \Lambda, \Psi, \Phi|X, m) \propto p(\Phi) |\Phi|^{-m_{\Psi} - m_{\Lambda}/2} |\Psi|^{-m_{\Psi} + m_{\Lambda}/2} |H|^{\nu/2} \exp \left( -\frac{1}{2} \text{trace} \left( \Psi^{-1}U \right) \right) \exp \left( -\frac{1}{2} \text{trace} \left( \Phi^{-1}FF' \right) \right) \quad (14) \]

where:

\[ U = (X - FA')\Phi^{-1}(X - FA') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B. \]

The conditional posterior density of the factor loadings given the factor scores, the disturbance covariance matrix, and the data is again normally distributed:

\[ p(\Lambda|F, \Psi, \Phi, X, m) \propto \exp \left( -\frac{1}{2} \text{tr} \left( \Psi^{-1}(\Lambda - \hat{\Lambda})(H + F\Phi^{-1}F')(\Lambda - \hat{\Lambda})' \right) \right) \quad (15) \]

where \( \hat{\Lambda} = (X'\Phi^{-1}F + \Lambda_0H)(H + F'\Phi^{-1}F)^{-1}. \) Comparing (15) with (7), we notice that the introduction of dependence between observations does affect the posterior distribution of factor loadings only for the incorporation of matrix \( \Phi \) in covariance matrix of \( F \). The conditional posterior density of the disturbance covariance matrix given the factor scores, the factor loadings, and the data is an inverted Wishart density:

\[ p(\Psi|F, \Lambda, \Phi, X, m) \propto |\Psi|^{-\nu/2} \exp \left( -\frac{1}{2} \text{tr} \left( \Psi^{-1}U \right) \right) \quad (16) \]

where \( U = (X - FA')\Phi^{-1}(X - FA') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B. \)
Finally, the conditional posterior distribution for the factor scores given the correlation matrix, the disturbance covariance matrix, the number of factors, the factor loadings and the data is normally distributed, with the following modification respect to (6):

\[
p(F|\Lambda, \Psi, \Phi, X, m) \propto \exp \left( -\frac{1}{2} tr \left( \Phi^{-1}(F - \tilde{F})(I_m + \Lambda'\Psi^{-1}\Lambda)(F - \tilde{F})' \right) \right),
\]

(17)

where \( \tilde{F} = X\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1} \).

We propose three different scenarios for \( \Phi \):

**First scenario** \( \Phi \) is a fix matrix, basically \( \Phi \) has two possible structures: either a conditional autoregressive structure: two regions are correlated if and only if they are adjacent regions. In alternative correlation between regions is inversely proportional to the distance between the region.

**Second scenario** \( \Phi \) has a prior inverted Wishart distribution with parameter \( G \):

\[
p(\Phi) \propto |\Phi|^{-\gamma/2} \exp \left( -\frac{1}{2} \text{trace} \ \Phi^{-1}G \right)
\]

(18)

We remove the assumption of independence between observations, but we do not specify any form of dependence and let the data determine that. For example, we can let \( G \) be a diagonal matrix, proportional to the identity matrix.

**Third scenario** We consider the same structures as in the first scenario, with the difference that the strength of correlation is a random variable, with a prior distribution itself. Basically we introduce another level on the hierarchical Bayesian model. Again, we consider two possible structures for \( \Phi \):

- \( \Phi \) has a conditional autoregressive structure, each region is correlated only with the adjacent regions. Furthermore we have to find (empirically) the range of \( \rho \) that guarantees \( \Phi \) to be a positive definite matrix and define a prior distribution for \( \rho \).
- Correlation between regions is inversely proportional to their Euclidean distance and the strength of correlation is a random variable itself. Basically, we assume \( \Phi = exp(-\rho D) \): \( exp \) indicates exponential of each single element, \( D \) is a fixed distances matrix and \( \rho \) is a random variable, whose distribution needs to be defined over a range of values that guarantees \( \Phi \) a positive definite matrix.

Once the Bayesian model is described, the assessment of the hyperparameters has to be considered. The process of hyperparameter assessment is described in Rowe (2003), while more sophisticated methods are proposed in Hayashi and Sen (2001).

We need to define the hyperparameters in equations (3)-(5). Lee and Press (1998) concluded that the estimation of \( \Lambda \) is not robust against change of values for hyperparameters \( \lambda_0 \). Maximum likelihood estimation from covariance method obtained with traditional factor analysis will be assumed as a prior mean for \( \Lambda \).
By definition, $H$ is any positive definite matrix, it is assumed that $H$ is a diagonal matrix or of the form $H = n_H I_m$, as a constant $n_H$ we choose the sample size. The same form is proposed for $B$, as $B = b_0 I_p$. The expected value of any diagonal element is

$$E(\psi_{ii}) = \frac{b_0}{\nu - 2p - 2}, \quad i = 1, \ldots, p.$$  

Substituting the training sample covariance matrix $\hat{\Sigma}$ and the a priori mean for the factor loadings in the above equations we have

$$\Psi_0 = \hat{\Sigma} - \Lambda_0 \Lambda_0',$$

then taking the average of the diagonal elements

$$\frac{1}{p} trace(\Psi_0) = \frac{1}{p} trace(\hat{\Sigma} - \Lambda_0 \Lambda_0'),$$

so

$$b_0 = \frac{n}{p} trace(\hat{\Sigma} - \Lambda_0 \Lambda_0').$$

Regarding the choice for the value of the hyperparameter $\nu$ we follow Rowe (2003), and define $\nu = n + 2 \times m + 2$.

4. Computations

In the previous section posterior distributions are found analytically, and Gibbs Sampling algorithm is implemented to generate samples. For Gibbs estimation of the posterior, we start with initial values for $F$ and $\Psi$, for example $\tilde{F}_{(0)}$ and $\tilde{\Psi}_{(0)}$. Remembering (15)-(17), leaving apart from now $\Phi$, then the cycle goes through:

$$\hat{\Lambda}_{i+1} = \text{a random sample from } P(\Lambda | \tilde{F}_i, \tilde{\Psi}_{(i)}, X)$$

$$\tilde{\Psi}_{i+1} = \text{a random sample from } P(\Psi | \tilde{F}_i, \hat{\Lambda}_{i+1}, X)$$

$$\tilde{F}_{i+1} = \text{a random sample from } P(F | \hat{\Lambda}_{i+1}, \tilde{\Psi}_{(i+1)}, X)$$

Finally, the means of the random sampling are the sampling based posterior marginal mean estimates of the parameters.

Concerning the sampling from posterior distribution of $\Phi$, we will show how to behave in the three situations mentioned before.

**First scenario** The first scenario represents on one hand the easiest form a computational point of view, on the other hand the weight of the prior structure is very strong.

**Second scenario** The second prior corresponds to the conjugate prior, so $\Phi$ has a posterior inverted Wishart distribution with updated parameters, more precisely:

$$E(\Phi | F, \Lambda, \Psi, X) \propto (X - F \Lambda')\Psi^{-1}(X - F \Lambda')' + FR^{-1}F' + D, \quad (19)$$

and degrees of freedom $(\gamma + p + m)$.

**Third scenario** Unfortunately we cannot solve analytically posterior distribution for $\rho$, furthermore posterior distribution is not log concave and sampling need implementation of adaptive rejection Metropolis sampling following Gilks et al. (1995).
5. Application

The Scotland lip cancer example is well-known in the literature and has been analyzed several times, since Clayton and Kaldor (1987) introduced it to study spatial distribution of cancer standardized mortality rates. In Scotland, there are 56 counties, and for each county the number of observed and expected cases for lips, lung, oral and breast cancer is given (expected numbers based on the population and its age and sex distribution in the county).

There is a slight evidence that social class has some effect on mortality for many types of cancer. Death from lung cancer, for example, occurs proportionately much more commonly among members of social classes consisting of unskilled or partially skilled workers. Increased exposure to sunlight has been implicated in the excess occurrence of lip cancers among rural populations, producing high rates in people who work outdoor. For each county, we measure one covariate measuring the percentage of the population engaged in agriculture, fishing, and forestry. This covariate is suspected to be related to sunlight exposure. The area close to the sea tends to be itself more rural area, and to has a higher percentage of population involved in outdoor work. For each county the proportion of all economically active households in the lowest social classes (class IV and V, partially skilled and unskilled). Density is also considered as a proxy of urban or rural area.

On the top of Figure 1 the observed standardized mortality ratios for lip and lung cancer for males in Scotland are shown. At the bottom of the same Figure, it is possible to observe, on the left, the distribution of proportion of people involved in outdoor activity, and, on the right, percentage of population belonging to lowest social class (as a proxy of socio-economic status). The data consists on a matrix $56 \times 7$; in Figure 2 the corresponding between observations sample correlation matrix is shown. The cell occupying the position $h, k$ is the $\text{cor}(x^h, x^k)$, the correlation between $h$ and $k$ countries. Observing the figure, the lack of independence between observations is evident.

Considering the second scenario, four latent variables (factor scores) were estimated by Bayesian factor analysis. Factor loadings are represented in Figure 3. The first factor is mostly determined by lips cancer and variable %agri indicating percentage of the population engaged in agriculture, fishing, and forestry, and variable density that is a proxy of the previous one. The first factor explains 67% of the entire variance. The second factor is mostly determined, on one hand, by socio-economic status, on the other hand by oral and lung cancer that are the cancer sites more related to socio-economic status. The third factor is mainly determined by cancer sites. The second and the third factor explain respectively 22% and 9% of the entire variation. The factor loadings are not sensitive to the prior distribution chosen for $\Phi$, on the other side the percent variation explained is. In Figure 4 factor scores are represented by geographical representation. First and third factor a strong spatial correlation is showed.

Interpretation of data reduction is essentially the same after assuming prior distribution as in third scenario. In particular, we assume $\Phi = \exp(-\rho D)$: $\exp$ indicates exponential of each single element, $D$ is a fixed distances matrix. In Figure 5, $D$ is represented, for each pair of region we calculate the minimum distance. $\rho$ is a random variable, with a uniform prior distribution over the range $[0.13, 1]$, to be sure $\Phi$ a positive definite matrix. In Figure 6, the posterior empirical distribution for $\rho$ it is showed, we obtain a posterior mean equal to 0.28. In Figure 7, posterior estimate for $\Phi$, as a function of $\rho$ is shown.
The Figure confirms the necessity of a spatial correlation and at the same time suggest to investigate data set where spatial correlation is stronger.

6. Conclusions

Although there has been a growing interest in the latent factor modeling of environmental and epidemiological data, the literature has provided few instruments for model building in presence of correlated data. The scenario we investigated is the presence of spatial dependence in the observed variables and latent factors.

Formal Bayesian statistical methods not only incorporate available prior information either from experts or previous data, but they allow the knowledge in these and subsequent data to accumulate in the determination of the parameter values. In the non-Bayesian Factor Analysis model, the factor loading matrix is determinate up to an orthogonal rotation. Typically after a non-Bayesian Factor Analysis, an orthogonal rotation is performed on the factor loading matrix according to one of many subjective criteria. This is not the case in Bayesian Factor Analysis. The rotation is automatically found. There is an entire probability distribution for the factor loading matrix and we determine its value statistically.

An important issue to be faced is the determination of the number of factors. We select the number of factors by empirical methods as percent variation: the resulting chosen number of factors is the minimum number that accounts for at least that amount of total variation in the observed covariance matrix. We compare our results with the ones obtained through a Bayesian approach. Defining $p(m)$, a prior on $m$, easily by Bayes’ Rule it is possible to compute the probability of each of the number of factors given the parameters

$$p(m|\mu, \Lambda, F, \Psi, X) \sim p(m)p(\mu)p(\Lambda|\Psi, m)p(F|m)p(\Psi)p(X|\mu, \Lambda, F, \Psi, m)$$

and determine the number of factors as the most probable. In this case, the results of the two approaches coincide, so we did not investigate further the latter mentioned method, although a probabilistic approach to the determination of the number of factors deserves to be better developed.

The proposed methods can be generalized to more complex models. In particular, a temporal structure can be introduced together with the spatial structure. To this aim, since the flexibility of the Bayesian approach allows us to define different structures for $\Phi$, the expression $\Phi = \Phi_T \otimes \Phi_S$, where $\Phi_T$ is defined as in Mezzetti and Billari (2205) for panel data, could be investigated.

References


**Figure 1:** Distribution of cancer standardized mortality ratio and risk factors in Scotland.

- **SMR lips cancer**
- **SMR lung cancer**
- **% employed in agriculture...**
- **% worker class IV and V**
Figure 2: Sample between observation correlation matrix.
Figure 3: Factor loadings.
Figure 4: Geographic representation of the four factor scores.
Figure 5: Matrix of minimum distance between each pair of region.
Figure 6: Posterior distribution for $\rho$. 
Figure 7: Posterior distribution of $\Phi$. 

![Posterior distribution of $\Phi$.](image)