SMA

Bayesian correlated factor analysis of socio-demographic indicators

Maura Mezzetti, Francesco C. Billari

Institute of Quantitative Methods and IGIER, Università Bocconi, Milano, Italy

Abstract. Recent changes in European family dynamics are often linked to common latent trends of economic and ideational change. Using Bayesian factor analysis, we extract three latent variables from eight socio-demographic indicators related to family formation, dissolution, and gender system and collected on 19 European countries within four periods (1970, 1980, 1990, 1998). The flexibility of the Bayesian approach allows us to introduce an innovative temporal factor model, adding the temporal dimension to the traditional factorial analysis. The underlying structure of the Bayesian factor model proposed reflects our idea of an autoregressive pattern in the latent variables relative to adjacent time periods. The results we obtain are consistent with current interpretations in European demographic trends.

Key words: Bayesian inference, Factor analysis, Correlated factor loadings, European family dynamics, Demographic convergence

1. Introduction

During the last decades, Europe has experienced tremendous changes in family dynamics. Fertility has fallen in some areas to unprecedented levels, marriage has lost its centrality in most societies and relationships between genders and generations have been shaped in a different way. The idea of Second Demographic Transition, coined during the mid 1980s by Lesthaeghe and van de Kaa (Van de Kaa, 1987; Lesthaege, 1995) emphasizes the importance of common trends in ideational change pervading all European societies. During the same period, the process of European integration pursued economic convergence and monetary integration as a key objective. As a consequence, also from the economic point of view countries belonging to the European Union have progressively experienced common trends.

In this paper, we present a temporal statistical model for the analysis of changing cross-national patterns in key socio-demographic variables: family formation, dissolution and gender systems. As in other studies of cross-national demographic

data, we use factor analysis, since it allows explaining the correlations between a large set of variables in terms of a small number of underlying factors. We first look for the underlying factors separately for each of the four years considered (1970, 1980, 1990, 1998).

In terms of knowledge on European family dynamics, this paper has two principal aims. Firstly, the paper aims to identify countries with similar development and as well as inter-country heterogeneity. In particular, we discuss whether a standard geographical aggregation in three broad European regions (North, South and East) is the most appropriate one. Secondly, once the different and similar patterns are identified, we seek an answer to the question of socio-demographic convergence among European countries. We aim to verify whether these changes have followed common trends towards convergence, as implied (at least in the long run) by the idea of Second Demographic Transition. Current patterns of diversity could in fact be accounted for by different rates at which various society are moving. The assumption of convergence follows from at least two considerations: firstly that socio-economics characteristics and ways of life have become similar across European countries, secondly that demographic behavior depends upon such factors (Coleman, 2002).

Common trends, according to the proponents of the Second Demographic Transition, cannot be singled out by looking at a single, specific indicator. Rather, they have to be detected by looking at the latent dimension(s) of social and demographic change. The literature on the topic has so far used techniques based on a frequentist approach for the reduction of macro-level social and demographic indicators (Pinelli, 2001).

Most proposals in the factor analysis literature assume that the data represent random, independent samples from a multivariate distribution (Lawley, 1940). 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. While Pinelli (2001) added a temporal dimension to the classical factorial analysis through a frequentist approach, our idea is that a Bayesian approach allows more flexibility to incorporate prior information about patterns in modernization process.

The innovation in this paper is that each country in each period represents an observation: each country is repeated four times. The methods introduced has, thus, to handle the dependence between the observations (Press and Shigemasu, 1997; Rowe, 1998). The idea behind the method proposed is to extend Rowe (1998) to multiple parallel time series and to estimate model parameters including factor scores that account for a temporal dependence. The flexibility of the Bayesian approach allows us to incorporate different temporal patterns for different European regions. The variables examined are thus a linear combination of the latent factors, and, therefore, have the same temporal pattern. A full description of the data set is present in Sect. 2. A detailed description of method proposed by Rowe (1998) is present in Sect. 3, and in Sect. 4 our proposal is shown, with a subsection aiming at the assessment of hyperparameters. After an illustration of the computational aspects in Sect. 5, results will be shown in Sect. 6. Sect. 7 contains final conclusions and remarks.

2. Data description and motivation

The data set we use has been collected by Paola di Giulio (whom we warmly thank for providing us with the data) and Antonella Pinnelli at the Department of Demographic Sciences of the University of Rome (Pinelli, 2001). The data set includes socio-demographic indicators that are related to the gender system, modernization, and family formation and dissolution. More specifically, the gender system is measured through: the percentage of seats in parliament occupied by women, as an indicator of women's participation in political decision making; women's activity rates as an indicator of participation in the labor market; the average age of women at first marriage as an indicator of the centrality of marriage in women's lives. Other demographic indicators are related to the quality of life (life expectancy at birth) or gender differentials in the quality of life (difference in life expectancy between women and men). In terms of family and fertility behavior, the data set contains the total fertility rate (average number of children per woman), the total divorce rate (an indicator of the prevalence of marital dissolution), and the percentage of births outside marriage, an indicator of disconnection between childbearing and marriage that is typically taken as the main indicator of demographic change in Europe following the Second Demographic Transition.

The information available concerns 19 European countries (of which most are members of the European Union) : Austria, Belgium, Bulgaria, Denmark, Finland, France, Greece, Hungary, Ireland, Italy, The Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland and United Kingdom. We thus have 8 variables for each year (1970, 1980, 1990 and 1998), a total of 76 observations on 8 variables (chosen out of the eleven variables present in the original data set).

To motivate the introduction of an innovative method to explain the temporal pattern of multidimensional data, we chose two out of eight variables to be briefly described more in detail: total fertility rates and the percentage of extra-marital births. The percentage of extra-marital births, besides being as we noticed one of the main indicators related to the Second Demographic Transition, is the variable with the highest coefficient of variation. The total fertility rate is, as known, a key variable for long-term population dynamics.

In Fig. 1 we show total fertility rates for each period and each country. The country is indicated with the initial (with minor exceptions: Pr stays for Portugal, since Po stays for Poland, and S stays for Switzerland since Sw stays for Sweden), and each color represents a different year. From 1970 to 1980 we observe a decline in fertility rates in all countries besides Poland. Fertility started to decline, especially in Western Europe, immediately after the baby boom, which took place in the first half of the sixties and involved all European countries but not Central and Eastern Europe. The decline was pronounced in all countries until 1975, reducing differences between countries. From 1975 to 1980, in the countries of Northern and Western Europe the total fertility rate has remained approximately constant. In 1970, the countries with highest fertility were Ireland with almost 4 children per woman, and Portugal and Spain with 2.9. In 1980, differences are less evident. After an increase in fertility rate from 1980 to 1990, in Denmark, Netherlands, Sweden and in Switzerland, the highest levels of fertility are observed in Northern Europe
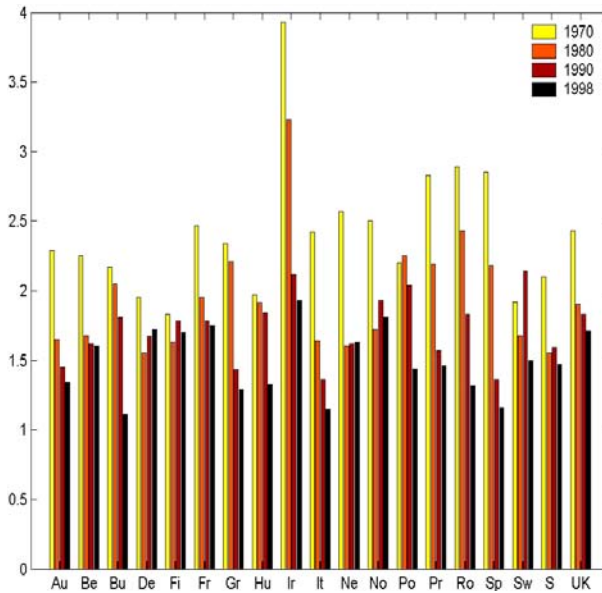
Fig. 1. Fertility rate in Europe from 1970 to 1998

(Sweden 2.1) and the lowest in Southern country (1.4 in Italy and Spain). Italy and Spain were subsequently the first countries to reach levels of lowest-low fertility–that is, under 1.3 children per woman. In 1970, total fertility for most Central and Eastern countries was higher than that in other countries. Levels were not homogenous between countries. In general, the large differences among countries of 1970 have decreased over time.

In Fig. 2, we show the percentage of extra-marital births for all countries through the different periods considered. The levels increase over time for all countries (Poland and Romania being exceptions regarding first and second period, where a slight decrease is observed), as one would expect with the spread of the Second Demographic Transition. The biggest increase in extra-marital births in the period 1970-1980 is observed for Sweden and Denmark, followed by Finland and Norway. From 1980 to 1990, France and the UK (after Norway) have the biggest change, reaching respectively 11.5% and 13.1%. From 1990 to 1998, central European countries experience the largest change and Ireland reaches almost 15%. The coefficient of variation reaches the lowest value in the last period, with some indication of convergence between countries with respect to this indicator. Nevertheless, the difference between countries at the two extremes (Sweden on one hand and Greece on the other side) has increased over time. A part from Greece and Italy, which have in 1998 respectively 3.9% and 9.2% of extra-marital births, other countries reach more homogeneous levels.
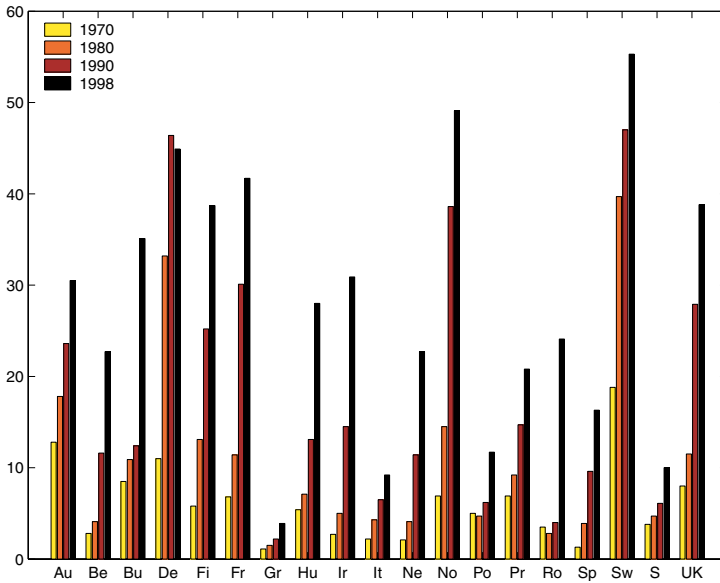
Fig. 2. Percentage of births outside marriage in Europe from 1970 to 1998

3. Correlated Bayesian factor model

We begin this section recalling classical and Bayesian factor analysis, in order to introduce the model developed in Rowe (1998). This will simplify the illustration of our approach in next section. We follow notation in Rowe (1998, 2003) and summarize his description of the model.

Factor analysis is used mainly in two situations. Sometimes it can be useful 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 smaller number of variables can be used to find a reasonable structure in the observed variables. This structure will aid in the interpretation and explanation of the process that has generated the observations. The second reason one would carry out a factor analysis is for data reduction. Since the observed variables are represented in terms of a smaller number of unobserved or latent variables, the number of variables in the analysis is reduced and so is the storage requirements.

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:

$$\begin{array}{ccccccccc} (x_i|\mu, \Lambda, f_i, m) & = & \mu & + & \Lambda & f_i & + & \varepsilon_i \\ (p \times 1) & & (p \times 1) & & (p \times m) & (m \times 1) & & (p \times 1) \end{array} \quad (1)$$

μ is a p-dimensional unobserved population mean vector,
Λ 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

ε_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 Ψ.

The parameters (μ, Λ, f, Ψ) 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. In Sect. 6 we will discuss the selection of the number of factors again. The estimate of the population mean μ is easily found by maximum likelihood and coincides with sample mean, see Lawley (1940). From now on, to simplify calculations but without loosing information, we will assume that x vector has zero mean, moreover if x is centered and scaled, then Λ is a correlation matrix between the x and f.

We can describe two kinds of models. In the first we can consider the factor scores as random vectors and in the second consider them as nonrandom vectors, which vary from one sample to another. Anderson and Rubin (1956) further show that the estimates of Λ and Ψ for random factor scores can be used for nonrandom factor scores in large samples due to asymptotic convergence. For these reasons, the model includes the factor scores as random quantities. As the model is overparameterized, the likelihood does not have a maximum, and we cannot reach maximum likelihood estimates through differentiating the log likelihood function.

To overcome this unidentifiability problem, we first assume that the factor scores are not fixed, but random normally distributed variables with mean 0, standard deviation 1 and correlation R, independent from error random variables ε_i. The variance and covariance matrix of observed vectors can be written as $Var(x_i|\Lambda, \Psi, m) = \Lambda R \Lambda' + \Psi$ and estimated by sample variance $\hat{\Sigma}$. After adding constraints on the parameters (as for example R being the identity and Λ being columnwise orthogonal), we can get unique solutions.

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.

We follow the approach by Press and Shigemasu (1997), since it can so far be considered the best and complete reference on this issue (Rowe, 1998). 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

$$p(X|F, \Lambda, \Psi, m) \propto |\Psi|^{-N/2} \exp\left(-\frac{1}{2}\text{trace}\left((X - F\Lambda')\Psi^{-1}(X - F\Lambda')'\right)\right)$$
(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 (Rowe, 2000a,b). The factor loadings are assumed to depend on the disturbance covariance matrix [using generalized conjugate priors, the previous dependence is removed (Rowe, 2001)]. 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:

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

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

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

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:

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

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:

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

where $\tilde{\Lambda} = (X'F + \Lambda_0 H)(H + F'F)^{-1}$.

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

$$p(\Psi|F, \Lambda, X, m) \propto |\Psi|^{-\frac{N+m+\nu}{2}} \exp\left(-\frac{1}{2}tr\left(\Psi^{-1}U\right)\right) \quad (8)$$

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:

$$
\begin{array}{ccccccc}
(x|\mu, \Lambda, f, m) & = & I_N \otimes \Lambda & f & + & \varepsilon & \\
(Np \times 1) & & (N \times N \otimes p \times m) & (Nm \times 1) & & (Np \times 1) &
\end{array} \quad (9)
$$

assuming the error has the following distribution:

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

the innovative idea proposed by Rowe (Rowe 1998, 2003) is to assume separable covariance matrix, that is $\Omega = \Phi \otimes \Psi$ ($\otimes$ indicate Kronecker product). If we let Φ be the identity matrix, we have the model in (1). In this way, $var(x_i|\Phi, \Psi, m.f, \Lambda) = \phi_{ii}\Psi$ 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.

4. Model proposed

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

$$
\begin{aligned}
p(X|F, \Lambda, \Psi, \Phi, m) \propto & |\Phi|^{-p/2}|\Psi|^{-N/2} \\
& \times \exp\left(-\frac{1}{2}\text{trace}\left(\Psi^{-1}(X - F\Lambda')'\Phi^{-1}(X - F\Lambda')\right)\right),
\end{aligned}
\tag{10}
$$

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

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

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

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

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

Basically, prior distributions for Ψ and Λ 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 Φ.

The likelihood in (10) requires the effort to define an appropriate prior distribution for the matrix Φ, that, hopefully, leaves the posterior analytically tractable. We propose three different prior distributions for the matrix Φ, starting from Rowe (1998). 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:

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

where:

$$U = (X - F\Lambda')'\Phi^{-1}(X - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B.$$

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

$$p(\Lambda|F, \Psi, \Phi, X, m) \propto \exp\left(-\frac{1}{2}tr\left(\Psi^{-1}(\Lambda - \tilde{\Lambda})(H + F'\Phi^{-1}F)(\Lambda - \tilde{\Lambda})'\right)\right) \tag{15}$$

where $\tilde{\Lambda} = (X'\Phi^{-1}F + \Lambda_0 H)(H + F'\Phi^{-1}F)^{-1}$. 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 Φ 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:

$$p(\Psi|F, \Lambda, \Phi, X, m) \propto |\Psi|^{-\frac{N+m+\nu}{2}}\exp\left(-\frac{1}{2}tr\left(\Psi^{-1}U\right)\right) \tag{16}$$

where $U = (X - F\Lambda')'\Phi^{-1}(X - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

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):

$$p(F|\Lambda, \Psi, \Phi, X, m) \propto \exp\left(-\frac{1}{2}tr\left(\Phi^{-1}(F - \tilde{F})(I_m + \Lambda'\Psi^{-1}\Lambda)(F - \tilde{F})'\right)\right), \tag{17}$$

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

The previous conditional posterior distributions basically does not depend on the prior we specify for Φ, as long as the prior distribution for Φ does not depend on other parameters. We propose three different prior distributions for Φ:

First prior We assume Φ as known, more specifically we assume a first order autoregressive structure within each country. Different countries are uncorrelated. Let indicate each observation with x_{ij}, where i indicates countries and j time (i.e. $i \in \{$Austria, Belgium, ..., Italy,..., UK $\}$, and $j \in \{$1970, 1980, 1990, 1998$\}$),

$$cov(x_{ij}, x_{i'j'}) = \begin{cases} 0 & i \neq i' \\ 1 & i = i' \text{ and } j = j' \\ \rho^{|j-j'|} & \text{if } i = i' \end{cases}$$

Second prior Φ has a prior inverted Wishart distribution with parameter G:

$$p(\Phi) \propto |\Phi|^{-\gamma/2} \exp\left(-\frac{1}{2}\text{trace } \Phi^{-1}G\right) \tag{18}$$

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 prior We assume the same prior as the previous one, with the difference that ρ, correlation between the same countries at different period is a random variable, with a prior distribution itself. Let us indicate x_{ij}, the generic row of matrix X, where i indicates country and j indicates period, let us suppose the index i varies firstly, the matrix Φ has the following form:

$$\Phi = \begin{bmatrix} 1 & 0 & \dots & \rho & 0 & \dots & \rho^2 & 0 & \dots & \rho^3 & 0 & \dots \\ 0 & 1 & 0 & \dots & \rho & 0 & \dots & \rho^2 & 0 & \dots & \rho^3 & 0 \\ 0 & \dots & 1 & 0 & \dots & \rho & 0 & \dots & \rho^2 & 0 & \dots & \rho^3 \\ \rho & 0 & \dots & 1 & 0 & \dots & \rho & 0 & \dots & \rho^2 & 0 & \dots \\ 0 & \rho & 0 & \dots & 1 & 0 & \dots & \rho & 0 & \dots & \rho^2 & 0 \\ 0 & \dots & \rho & 0 & \dots & 1 & 0 & \dots & \rho & 0 & \dots & \rho^2 \\ \rho^2 & 0 & \dots & \rho & 0 & \dots & 1 & 0 & \dots & \rho & 0 & \dots \\ 0 & \rho^2 & 0 & \dots & \rho & 0 & \dots & 1 & 0 & \dots & \rho & 0 \\ \dots & 0 & \rho^2 & 0 & \dots & \rho & 0 & \dots & 1 & 0 & \dots & \rho \\ \rho^3 & \dots & 0 & \rho^2 & 0 & \dots & \rho & 0 & \dots & 1 & 0 & \dots \\ 0 & \rho^3 & \dots & 0 & \rho^2 & 0 & \dots & \rho & 0 & \dots & 1 & 0 \\ 0 & \dots & \rho^3 & \dots & 0 & \rho^2 & 0 & \dots & \rho & 0 & \dots & 1 \end{bmatrix} \tag{19}$$

Let Φ_T be the matrix for the AR(1) model with four time points

$$\Phi_T = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \tag{20}$$

Φ can be written as

$$\Phi = \Phi_T \otimes I_{19}. \tag{21}$$

The observations for the same country but for different time points are correlated according to a first order Markov scheme. Different structure for previous matrix can be given, for example ρ can vary for the different countries. Our proposal is a Beta prior distribution for ρ

4.1. Hyperparameters assessment

Finally, when 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 Eqs. (3)–(5). Lee and Press (1998) concluded that the estimation of Λ is not robust against change of values for hyperparameters Λ_0. Following Rowe (2003), our proposal is using as a prior maximum likelihood estimation from covariance method obtained with traditional factor analysis, leaving the prior precision quite small.

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

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

so

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

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

5. 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 Ψ, for example $\tilde{F}_{(0)}$ and $\tilde{\Psi}_{(0)}$. Remembering (15)-(17), leaving apart from now Φ, then the cycle goes through:

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

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 Φ, we will show how to behave in the three situations mentioned before.

First prior The first prior is the easiest form a computational point of view, but it requires to be very confident about temporal structure of data.

Second prior The second prior corresponds to the conjugate prior, so Φ has a posterior inverted Wishart distribution with updated parameters, more precisely:

$$E(\Phi|F, \Lambda, \Psi, X) \text{ proportional to}$$
$$(X - F\Lambda')\Psi^{-1}(X - F\Lambda')' + FR^{-1}F' + D, \tag{22}$$

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

Third prior From Eq. (21), we can derive the determinant and inverse of Φ as a function of ρ;

$$|\Phi| = (1 - \rho^2)^{(n_T-1)\times n_C},$$

where n_t is the number of time knots and n_C is the number of countries.

$$\Phi^{-1} = \Phi_T^{-1} \otimes I_{n_C},$$

where

$$\Phi_T^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & -\rho & 1 \end{bmatrix} \tag{23}$$

Letting ρ to be a random variable, a Beta distribution seems a natural choice as a prior distribution for ρ. Different choices of hyperparameters were investigated, in particular starting from sample covariance matrix $\hat{\Phi}$, reasonable value for prior expected value and prior precision are proposed. We obtain the following posterior distribution, starting from results in Rowe (2003):

$$p(\rho|F, \Lambda, \Psi, m, X) \sim p(\rho)(1 - \rho^2)^{-\frac{(N-1)(p+m)}{2}} \exp\left(-\frac{k_1 - k_2\rho + k_3\rho^2}{2(1-\rho^2)}\right) \tag{24}$$

Let $C = (X - F\Lambda')\Psi^{-1}(X - F\Lambda')' + FR^{-1}F'$ and $k_1 = \text{trace}(C)$, $k_2 = \text{trace}(M_2C)$ and $k_3 = \text{trace}(M_3C)$, M_2 and M_3 are matrices of zeros and ones appropriately defined: $M_2 = M_{02} \otimes I_{19}$,

$$M_{02} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$M_3 = M_{03} \otimes I_{19}$,

$$M_{03} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

An adaptive rejection sampling algorithm is written to sample from (24).

6. Results

In Fig. 3, the sample matrix Φ is shown. The cell occupying the position h, k is the $cor(x'_h, x'_k)$, the correlation between h and k countries. Since the elements represent countries in different time periods, the generic h element is the relative to in j^{th} period, where j is the nearest integer less or equal to h divided by number of countries plus one, and i^{th} country, where $i = h - j \times n_C$, where n_C stays for number of countries.
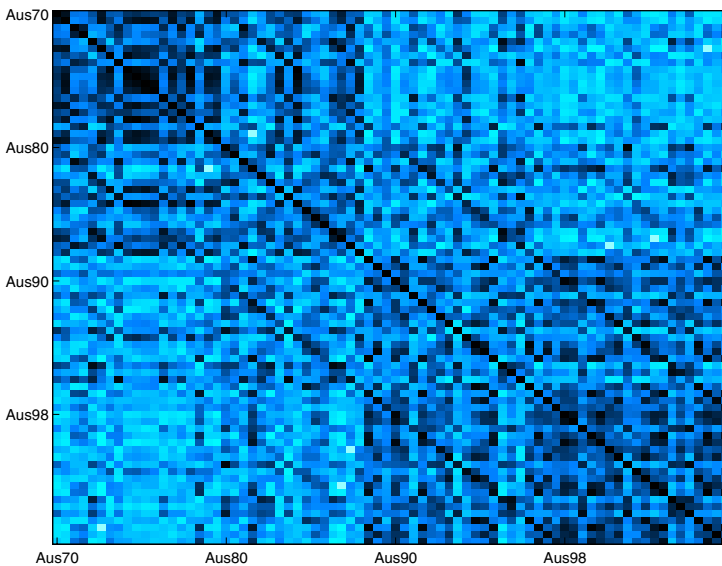

Fig. 3. Sample Φ

Lighter colors represent lower covariance, and darker colors represent higher covariance. As expected, the principal diagonal is black (being the elements equal to 1). Observing the figure, we notice that the diagonal starting from element $\{n_C + 1, 1\}$ (and symmetrically the one starting in $\{1, n_C + 1\}$) is quite dark, indicating the presence of correlation between observations related to the same country in adjacent time periods. Less marked, but still visible, is the diagonal starting in $\{2 \times n_C + 1, 1\}$, indicating the presence of correlation between observations related to the same country at a time distance of two periods.

We can notice some darker areas even outside the previous described diagonals, indicating the high correlation between data of different countries. In particular, in the first period high correlation is observed among countries in Western Europe marked by a prevalence of the Catholic religion (Italy, Ireland, Spain and Portugal). In contrast, in the last period (last block around principal diagonal), high correlation is present among Nordic countries (Sweden and Norway on one side and Denmark and Finland on the other side). Three latent variables (factor scores) were estimated by Bayesian factor analysis. Factor loadings are represented in Fig. 4. The first factor is mostly determined by percentage of extra-marital births, total divorce rates, percentage of parliamentary seats held by women and women's activity rates. The first factor explains 70% of the entire variance. Extreme values for first factor are observed in Scandinavian countries on one hand, and southern countries together with Ireland on the other side (Fig. 5). This factor is clearly related with modernization and ideational change, as depicted in the idea of Second
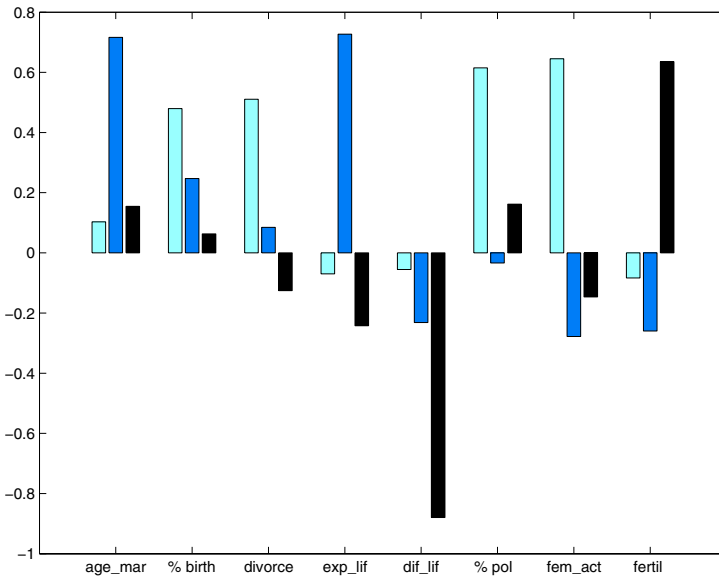

Fig. 4. Factor loadings. *age-mar* stays for age at first marriage, % *birth* stays for percentage births outside marriage, *divorce* indicates divorce rate , *exp-lif* stays for life expectation, *dif-life* difference in life expectation between males and females, % *pol* percentage of sets in parliament occupied by women, *fem-act* percentage of women active and, finally, *fertil* is fertility rate
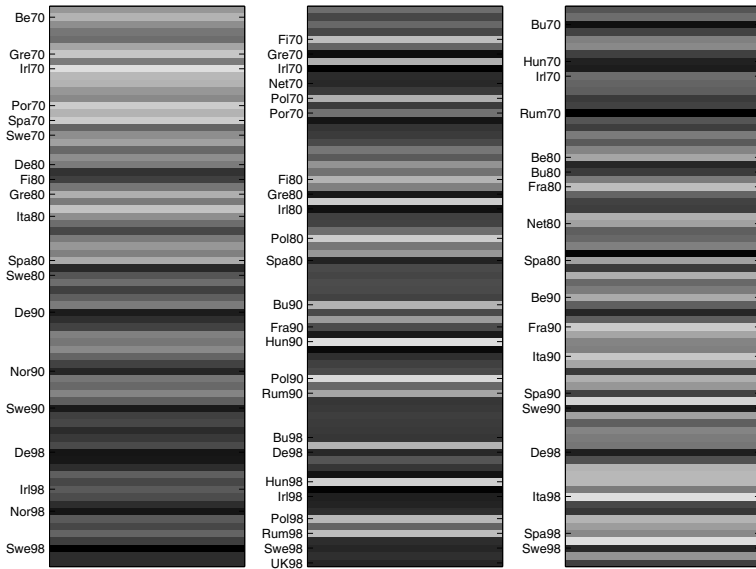
Fig. 5. Factor scores

Demographic Transition. In Fig. 6, the first factor relative to 1990 is projected over map of Europe, to clarify the contrast between countries with high value on the first factor score and countries with low value on the first factor score. The second factor is mostly determined by the age at first marriage and life expectancy; it is inversely related with total fertility rates and women's activity rates. The second factor thus seems to be affected by economic conditions, contrasting on one side richer countries (with better health condition) with on the other side poorer countries (with women's activity rates being traditionally high in Central and Eastern Europe). Higher values on this factor are reached in Northern Europe and lower values in eastern Europe. In Fig. 7, the second factor relative to 1998 is projected over map of Europe. The third factor selected is mostly determined with negative sign by the difference in life expectancy between genders, and it is also positively associated with total fertility rates. As shown in Fig. 5, lower values are reached for Southern European countries, while higher values for Eastern countries at the beginning of the study and Northern countries in the last periods. The interpretation of this factor is much less clear than it is for the first two factors. In Fig. 8, third factor relative to 1998 is projected over map of Europe.

An important issue to be faced is the determination of 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

$$p(m|\mu, \Lambda, F, \Psi, X) \sim p(m)p(\mu)p(\Lambda|\Psi, m)p(F|m)p(\Psi)p(X|, \Lambda, F, \Psi, m)$$
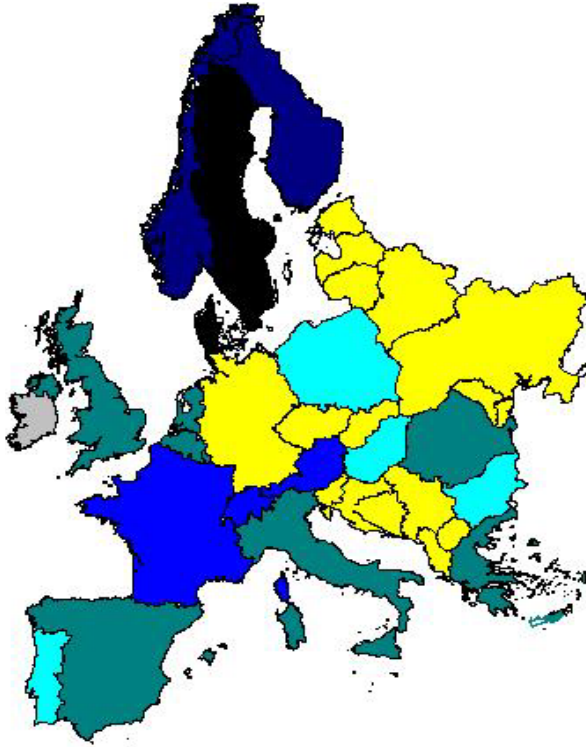
Fig. 6. Projection of first factor over 1990. Yellows indicated countries were information were not available. Colors vary between weak gray (corresponding to Ireland, with a value equal -0.60), light blue (corresponding to Eastern Europe, reaching -0.30), finishing to black corresponding to Sweden with 2.05

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 determination of number of factors deserves to be better developed.

7. Conclusions

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.
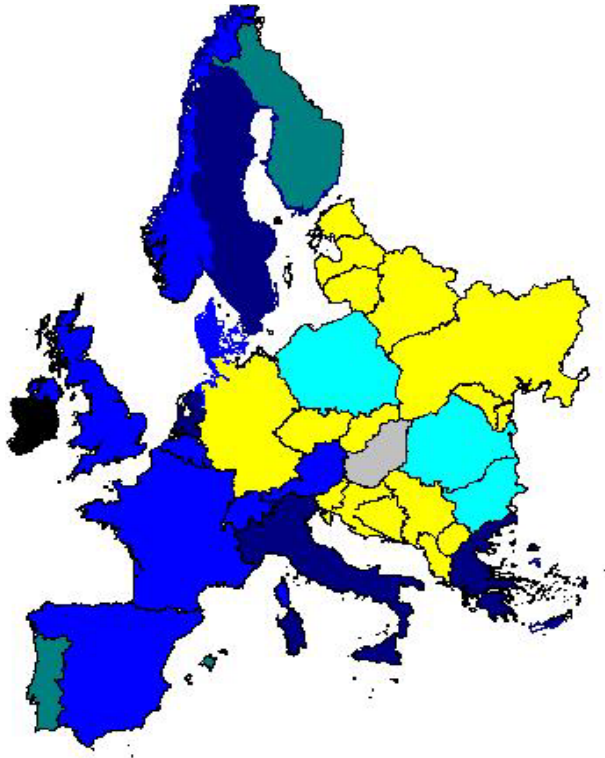
Fig. 7. Projection of second factor over 1998. Yellows indicated countries were information were not available. Colors vary between weak gray (corresponding to Hungary, with a value equal -1.11), light blue (corresponding to Eastern Europe, reaching -0.91), finishing to black corresponding to Ireland with 1.77

In this paper, we have shown that the flexibility of the Bayesian approach allows us to introduce an innovative temporal factor model, extending the temporal dimension in factor analysis to multiple parallel observations. The underlying structure of the temporal factor model proposed reflects our idea of an autoregressive pattern in the latent variables addressed relative to adjacent time periods. The results we obtained, with the first latent variable explaining a large share of variance and connected to modernization, are consistent with current interpretations in European demographic trends.

The methods proposed can be easily generalized to more complex models, as for example, introduction of a spatial structure together with the temporal one. The flexibility of the Bayesian approach allows us to define different structure for Φ, an expression to be investigated is the definition of $\Phi = \Phi_T \otimes \Phi_S$, where Φ_S is an adequate matrix defined in order to express the correlation between adjacent countries.
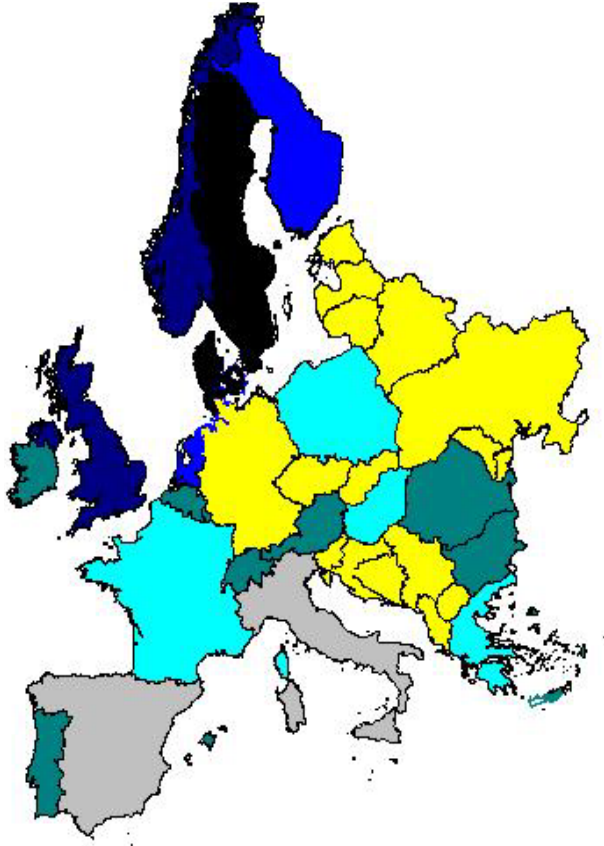
Fig. 8. Projection of third factor over 1998. Yellows indicated countries were information were not available. Colors vary between weak gray (corresponding to Italy and Spain, with a value equal -1.24), light blue (corresponding to Eastern Europe, reaching -0.90), finishing to black corresponding to Sweden with 1.61

References

Anderson TW, Rubin H (1956) Statistical inference in factor analysis. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, edited by Jerzy Neyman, 111–150

Coleman DA (2002) Populations of the industrial world: A convergent demographic community. International Journal of Population Geography, 8: 319–344

Hayashi K, Sen PK (2001) Bias-Corrected estimator of factor loadings in bayesian factor analysis. Educational and Psychological Measurement, 62(6): 944–959

Haskey J (2000) Demographic Issues in 1975 and 2000. Population Trends, 100: 20–31

Lawley DN (1940) The estimation of factor loadings by the method of maximum likelihood Proceedings of the Royal Society of Edimburgh, 60: 84–82

Lee SE, Press SJ (1998) Robustness of Bayesian factor analys estimates. Communications in Statistics-Theory and Methods, 27: 1871–1893

Lesthaeghe R (1995) The second demographic transition in western countries: an interpretation. In: Oppenheim Mason K, Jensen AM (eds) Gender and family change in industrialized countries, pp: 17-62 Clarendon Press, Oxford

Pinnelli A, Hoffman-Nowotny HJ, Fux B (2001) Fertility and new types of households and family formation in Europe Council of Europe Publishing, Population Studies N. 35, Strasbourg

Press SJ, Shigemasu K (1997) Bayesian inference in factor analysis-revised, with an appendix by Rowe DB Technical Report No. 243, Department of Statistics, University of California, Riverside, CA

Rowe DB (1998) Correlated bayesian factor analysis Thesis PhD Department of Statistics, University of California, Riverside, CA

Rowe DB, Press SJ (1998) Gibbs sampling and hill climbing in bayesian factor analysis Technical Report No. 255, Department of Statistics, University of California, Riverside, CA

Rowe DB (2000a) A Bayesian factor analysis model with generalized prior information. Social Science Working Paper 1099, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA

Rowe, D. B. (2000b) Incorporating prior knowlege regarding the mean in bayesian factor analysis. Social Science Working Paper 1097, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA

Rowe DB (2001) A bayesian model to incorporate jointly distributed generalized prior information on means and loadings in factor analysis. Social Science Working Paper 1110, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA

Rowe DB (2003) Multivariate bayesian statistics: models for source separation and signal unmixing. CRC Press, Boca Raton, FL, USA

Van de Kaa D (1987) Europe's second demographic transition Population Bulletin, 42(1)