CHAPTER 5

MODERN BAYESIAN FACTOR ANALYSIS

HEDIBERT FREITAS LOPES

George Washington University, D.C., USA

5.1 Introduction

The origin of factor analysis can be traced back to Spearman's (1904) seminal paper on general intelligence. At the time, psychologists were trying to define intelligence by a single, all-encompassing unobservable entity, the g factor. Spearman studied the influence of the g factor on examinees' test scores on several domains: pitch, light, weight, classics, French, English, and mathematics. At the end of the day, the g factor would provide a mechanism to detect common correlations among such imperfect measurements. More precisely, Spearman's (1904) one-factor model based on p test domains (measurements) and n examinees (individuals) can be written as

$$y_{ij} = \mu_j + \beta_j g_i + \varepsilon_{ij}, \tag{5.1}$$

Bayesian Inference in the Social Sciences.

for $i = 1, \ldots, n$, $j = 1, \ldots, p$, where y_{ij} is the score of examinee i on test domain j, μ_j is the mean of test domain j, g_i is the value of the intelligence factor for person i, β_j is the loading of test domain j onto the intelligence factor g, and ϵ_{ij} is the random error term for person i and test domain j.

Spearman spends part of his 90-page paper defending his one-factor model of general intelligence, arguably his main, seminal contribution to the fields of psychometrics as well as statistical modeling at this time. Nonetheless, Bartholomew's (1995) review paper starts by stating that "*Spearman invented factor analysis but his almost exclusive concern with the notion of a general factor prevented him from realizing its full potential.*"

For subsequent developments, mainly in psychology studies, see Burt (1940), Holzinger and Harman (1941), and Thomson (1953), amongst others, where the factors had an *a priori* known structure. The extension to multiple factors as well as its formal statistical framework came many decades later. Multiple factor analysis was first introduced by Thurstone (1935,1947) and Lawley (1940,1953), along with estimation via the centroid method and maximum likelihood, respectively. Hotelling (1955) proposed a more robust method of estimation, the method of principal components, while Anderson and Rubin (1956) formalized and elevated factor analysis to the realm of statistically and probabilistically sound modeling schemes.

Computationally speaking, maximum likelihood estimation became practical in the late 1960s through the work of Jöreskog (1967,1969). A further improvement was achieved in the early 1980s through the EM algorithms of Rubin and Thayer (1983,1983); see also Bentler and Tanaka (1983). In the late 1980s, Anderson and Amemiya (1988) and Amemiya and Anderson (1990) studied the asymptotic behavior of estimation and hypothesis testing for a large class of factor analysis under general conditions, while Akaike (1987) proposed an information criterion to selecting the proper number of common factors.

Table 5.1 illustrates the fast increase in the use of factor analysis in a few areas of science and industry over the last century. This chapter discusses the Bayesian contribution to this table, most of which appearing after the mid-1990s due to both the increasing access to faster, smaller, cheaper and sharable computers and processors and the revival and/or introduction of efficient Monte Carlo schemes for posterior inference in highly structured stochastic systems (Gamerman and Lopes, 2006), of which factor analysis is an increasingly popular member.

Factor Analysis at 100: To celebrate the centennial of Spearman (1904), The L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill, hosted in May 2004 a workshop entitled *Factor Analysis at 100: Historical Developments and Future Directions*, with David Bartholomew, Michael Browne, Robert Jennrich, Karl Jöreskog, Yasuo Amemiya and many other distinguished invited speakers. A thought-provoking historical account is the paper *Three Faces of Factor Analysis* presented at the meeting by

Table 5.1: Distribution of papers on factor analysis on the Internet. Reproduced from Kaplunovsky (2004), who presents more areas of science and industry and more disaggregated time intervals.

	1904-1985 80 years	1986-1995 10 years	1996-2004 10 years
Psychology	179	378	723
Psychiatry	29	100	236
Medicine	62	131	225
Spectroscopy	38	90	198
Chemistry	26	89	165
Biology	35	43	88
Physiology	46	77	80
Geriatry	14	19	56
Economics	26	11	46
Chromatography	11	38	39

Bartholomew. Another historical account was neatly organized by a group of students and faculty at the Thurstone Psychometric Lab entitled *Factor Analysis Genealogy* and *Factor Analysis Timeline*. See the conference webpage for these and additional information at `http://www.fa100.info`. The papers presented at the meeting appeared in Cudeck and MacCallum (2007).

The rest of the chapter is organized as follows. Section 5.2 reviews and explores the basic normal linear factor model, identification issues, estimation of parameters and factors as well as estimation/selection of the number of common factors. Section 5.3 deals with factor variance models, commonly present in financial econometrics contexts. Factor models for spatial and space-time problems are introduced in Section 5.4. Section 5.5 presents recent developments in factor analysis, such as prior and posterior robustness, mixture of factor analyzers, factor analysis in time series and macroeconometric modeling and sparse factor structures. Some of the recent contributions to the literature on non-Bayesian (large dimensional and/or dynamic) factor analysis are presented in Section 5.6. The chapter concludes with Section 5.7.

5.2 Normal Linear Factor Analysis

Let $y_i = (y_{i1}, \ldots, y_{ip})'$, for $i = 1, \ldots, n$, be a p-dimensional vector with the measurements on p related variables (Spearman's tests, attributes, macroeconomic or financial time series, census sectors, monitoring stations, to name a few examples). The basic normal linear factor model assumes that ys are independent and identically distributed $N(0, \Omega)$, i.e., a zero-mean multivariate normal with a $p \times p$ non-singular variance matrix Ω. Loosely speaking, a factor model usually rewrites Ω, which depends of $q = p(p+1)/2$ variance and covariance components, as a function of d parameters, where d is potentially many orders of magnitude smaller than q. Figure 5.1 illustrate a 3-factor model for a $p = 9$-dimensional vector of attributes.

More specifically, for any positive integer $k \leq p$, a standard normal linear k-factor model for y_i is written as

$$y_i|f_i, \beta, \Sigma \sim N(\beta f_i, \Sigma), \quad (5.2)$$

$$f_i|H \sim N(0, H), \quad (5.3)$$

where f_i is the k-dimensional vector of common factors, β is the $p \times k$ matrix of factor loadings, $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ is the covariance of the specific factors and $H = \text{diag}(h_1, \ldots, h_k)$ is the covariance matrix of the common factors. The uniquenesses σ_i^2s, also known as idiosyncratic or specific variances, measure the residual variability in each of the data variables once that contributed by the factors is accounted for. Conditionally on the common factors, f_k, the measurements in y_i are independent. In other words, the common factors explain all the dependence structure among the p variables and, based on equations (5.2) and (5.3), the unconditional, constrained covariance matrix of y_i becomes

$$\Omega = \beta H \beta' + \Sigma. \quad (5.4)$$

The matrix Ω depends on $d = (p+1)(k+1) - 1$, the number of elements of β, H and Σ, a number usually considerably smaller than $q = p(p+1)/2$, the number of elements of the unconstrained Ω.

5.2.1 Parsimony

In practical problems, especially with larger values of p, the number of factors k will often be small relative to p, so most of the variance-covariance structure is explained by a small number of common factors. For example, when $p = 100$ and $k = 10$, a configuration commonly found in modern applications of factor analysis, $q = 5050$ and $d = 1110$, or roughly $q = 5d$. Similarly, when $p = 1000$ and $k = 50$, if follows that $q = 500500$ and $d = 51050$, or roughly $q = 10d$. Such a drastic reduction in the number of unrestricted parameters renders factor modeling inherently a parsimony-inducing technique.

5.2.2 Identifiability

The k-factor model of equations (5.2) and (5.3) is invariant under transformations of the form $\beta^* = \beta P'$ and $f_i^* = Pf_i$, where P is any orthogonal $k \times k$ matrix. Likelihood identification is achieved by assuming that β is a block lower triangular matrix of rank k and with ones in the main diagonal. It is worth noting that the ones in the main diagonal of β can be replaced

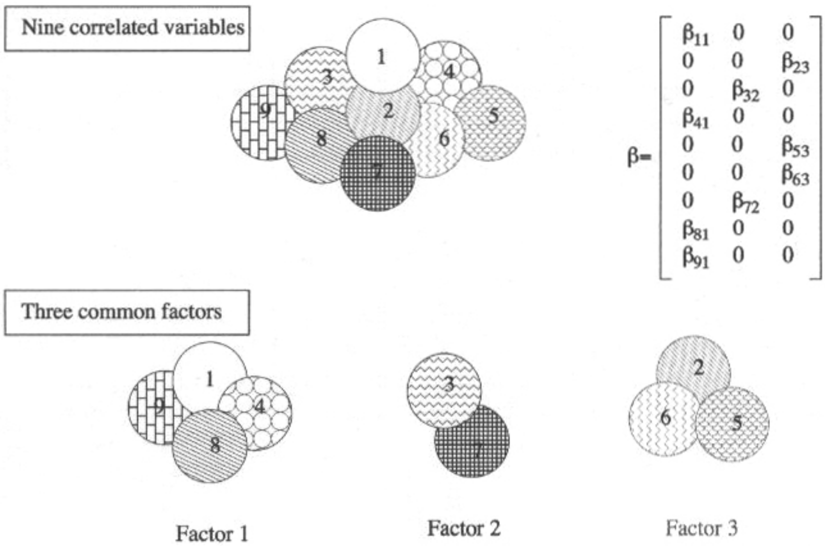

Figure 5.1: Illustration of a 3-factor model structure for $p = 9$ correlated variables. Notice that each variable is explained by only one of the $k = 3$ common factors. Therefore, $\Omega = \beta H \beta' + \Sigma$ is block-diagonal, with independent blocks defined by variables (y_1, y_4, y_8, y_9), (y_3, y_7) and (y_2, y_5, y_6).

by strictly positive values, as long as H, the common factors' covariance matrix, is replaced by I_k. This form is used, for example, in Geweke and Zhou (1996) and Lopes and West (2004), as well as the majority of the Bayesian factor modelers after them, and provides both identification and, often, useful interpretation of the factor model.

With Ω of full rank p, the resulting factor form of Ω has $d = p(k+1) - k(k-1)/2$ parameters, compared with the total $q = p(p+1)/2$ in an unconstrained (or $p = k$) model, leading to the constraint that

$$p(p+1)/2 - p(k+1) + k(k-1)/2 \geq 0, \tag{5.5}$$

which provides an upper bound on k. Even for small p, the bound will often not matter as relevant k values will not be so large. In realistic problems, with p in double digits or more, the resulting bound will rarely matter. Finally, note that the number of factors can be increased beyond such bounds by reducing the rank of Σ. See also Ihara and Kano (1995), who establish and discuss conditions for full, marginal and conditional model identification are discussed.

5.2.3 Invariance

Three important results for the class of standard normal linear factor models are shown below: *(i)* non-diagonal H is irrelevant, *(ii)* they are invariant to the order of the variables, and *(iii)* reduced rank factor loadings and non-identifiability.

Result (i): Let us assume that H is non-diagonal and that $H = LL'$. Then, equations (5.2) could be rewritten as $y \sim N(\beta^* f^*, \Sigma^*)$, where $\beta^* = \beta L$, $f^* = L^{-1} f$ and $\Sigma^* = \Sigma$ unchanged. The rotated common factors f^* are still zero-mean multivariate normal with covariance matrix $I_q = L^{-1} H (L^{-1})'$. The loading matrix β^* can be transformed into a lower block triangular matrix by letting $\beta^* = ((\beta_1^*)', (\beta_2^*)')'$, β_1^* a $(k \times k)$ matrix, β_2^* a $(p - k \times k)$ matrix. So,

$$\tilde{\beta} = \beta^*(Z')^{-1} = \begin{pmatrix} Z \\ \beta_2^*(\beta_1^*)'(Z')^{-1} \end{pmatrix} \quad \text{and} \quad \tilde{f} = Z'(\beta_1^*)^{-1} f^*,$$

where $\beta_1^*(\beta_1^*)' = ZZ'$, for lower triangular Z. More compactly, $\tilde{\beta} = \beta P_1$ and $\tilde{f} = P_2 f$ for $P_1 = L(Z')^{-1}$ and $P_2 = Z'(\beta_1^*)^{-1} L^{-1}$. Now, it is straightforward to see that

$$\begin{aligned} P_2 H P_2' &= Z'(\beta_1^*)^{-1} L^{-1} H (L^{-1})'((\beta_1^*)^{-1})' Z \\ &= Z'(\beta_1^*)^{-1}((\beta_1^*)^{-1})' Z = Z'(\beta_1^*(\beta_1^*)')^{-1} Z \\ &= Z'(ZZ')^{-1} Z = Z'(Z')^{-1} Z^{-1} Z = I_k, \end{aligned}$$

so equations (5.2) and (5.3) can be recovered by letting β be $\tilde{\beta}$ with column i normalized by $\tilde{\beta}_{ii}$, for $i = 1, \ldots, q$, and $H = \text{diag}(\tilde{\beta}_{11}^2, \ldots, \tilde{\beta}_{kk}^2)$.

Result (ii): It follows directly and similarly since reordering the components of y is equivalent to pre-multiplying it by a permutation matrix Q, accordingly pre-multiplying the factor loadings matrix by Q and post-multiplying it by L, i.e., $\beta^* = Q\beta L$, while $f^* = L^{-1} f$ and $\Sigma^* = Q\Sigma Q'$ are still diagonal. β^* is transformed into a lower block triangular matrix and f rotated to produce orthogonal factors by repeating the steps of result *(i)*.

Result (iii): It is important to emphasize that the above two results apply only when k is the right number of factors, i.e., when the rank of the loading matrix β is k. More precisely, when $H = I_k$, Geweke and Singleton (1980) showed that, if β has rank $r < k$, then there exists a matrix Q such that $\beta Q = 0$ and $Q'Q = I_k$ and, for any $(p \times k)$ orthogonal matrix M, it follows that

$$\beta\beta' + \Sigma = (\beta + MQ')'(\beta + MQ') + (\Sigma - MM'). \tag{5.6}$$

This translation invariance of Ω under the factor model implies a lack of identification and, in application, induces symmetries and potential multimodalities in the resulting likelihood functions. This issue relates directly to the question of uncertainty of the number of factors. Figure 5.3 of Example 5.1 illustrates this phenomenon. See also Lopes and West (2004) for further discussion and additional empirical evidence. In Frühwirth-Schnatter and Lopes (2009), we take advantage of the result *(iii)* and propose a new stochastic search scheme that navigates the joint space of the sparse factor loadings and the number of common factors.

5.2.4 Posterior Inference

Early MCMC-based posterior inference in standard factor analysis appears in, among others, Geweke and Zhou (1996) and Lopes and West (2004). They basically propose and implement a standard Gibbs sampler that cycles through the full conditional distributions of $p(\beta|f,\Sigma,y)$, $p(\Sigma|f,\beta,y)$ and $p(f|\beta,\Sigma,y)$, which following well known distributions when conditionally conjugate priors are used. Here we assume that $H = I_k$ and that β is block lower triangular, for identification, with diagonal strictly positive diagonal elements.

Prior Specification. The unconstrained components of β are independent and identically distributed (i.i.d.) $N(m_0, C_0)$, the diagonal components β_{ii}s are i.i.d. truncated normal from below at zero, denoted here by $N_{(0,\infty)}(m_0, C_0)$, and the idiosyncratic variances, σ_i^2 are i.i.d. $IG(\nu/2, \nu s^2/2)$. The hyperparameters m_0, C_0, ν and s^2 are known. It is worth mentioning that the above prior specification has been extended and modified many times over to accommodate specific characteristics of the scientific modeling under consideration. Lopes et al. (2008), for example, utilize spatial proximity to parameterize the columns of β when modeling pollutants across Eastern US monitoring stations. See the following sections for more structured factor loading matrices, as well as common and specific factor dynamics.

Full Conditional Distributions. The factor model can be seen as a standard multivariate regression model when deriving the full conditionals $(f|\beta,\sigma,y)$ and $(\beta|\Sigma,f,y)$ (Box and Tiao, 1973; Press, 1982; Broemeling, 1985; Zellner, 1971; Gamerman and Lopes, 2006). More precisely, let us start by denoting $y = (y_{(1)},\ldots,y_{(p)})$, $\beta = (\beta_1',\ldots,\beta_p')'$ and $F_i = (f_{(1)},\ldots,f_{(i)})$, with $\beta_i = (\tilde{\beta}_i', 0_{k-i})'$ and $F_i = (f_{(1)},\ldots,f_{(i)})$, for $i = 1,\ldots,k-1$. Therefore, for $i = 1,\ldots,n$, $(f_i|\beta,\Sigma,y) \sim N((I_k+\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}y_i, (I_k+\beta'\Sigma^{-1}\beta)^{-1})$. For $i = 1,\ldots,k$, $(\tilde{\beta}_i'|\Sigma,f,y) \sim N(m_i,C_i)\delta_{\beta_{ii}>0}$, where $m_i = C_i(C_0^{-1}\mu_0 1_i + \sigma_i^{-2}F_i'y_{(i)})$, $C_i^{-1} = C_0^{-1}I_i + \sigma_i^{-2}F_i'F_i$, 1_i is an i-dimensional vector of ones and δ_x is the indicator function at x. For $i = k+1,\ldots,p$, $(\beta_i|\Sigma,f,y) \sim N(m_i,C_i)$, where $m_i = C_i\left(C_0^{-1}\mu_0 1_k + \sigma_i^{-2}F_k'y_{(i)}\right)$ and $C_i^{-1} = C_0^{-1}I_k + \sigma_i^{-2}F_k'F_k$. Finally, $(\sigma_i^2|\beta,f,y) \sim IG((\nu+n)/2, (\nu s^2+d_i)/2)$, where $d_i = (y_{(i)}-F_k\beta_i')'(y_{(i)}-F_k\beta_i')$.

Chib et al. (2006), in the context of multivariate stochastic volatility modeling, propose an MCMC scheme that jointly samples the factor loadings matrix β and the common factor scores f_is. They basically integrate the factors out of the full conditional distribution of β and use a Metropolis-Hastings step to sample β. Then they sample f given β, as described in our scheme above. See also Ghosh and Dunson (2009) for a slight modification of our scheme, and Song and Lee (2001) for a factor analysis combining continuous and polytomous data. Paisley and Carin (ICML) propose a nonparametric extension to the factor analysis problem using a beta process prior.

EXAMPLE 5.1

Lopes and West (2004) fit one-, two- and three-factor models to monthly international exchange rate data spanning from January of 1975 to December 1986 (a total of $n = 144$ observations). The time series are the exchange rates in British pounds of the US dollar (US), Canadian dollar (CAN), Japanese yen (JAP), French franc (FRA), Italian lira (ITA) and (West) German (Deutsch)mark (GER). Here we replicate the setup used in Lopes and West (2004), with prior hyperparameters $m_0 = 0$, $C_0 = 1$, $\nu = 2.2$ and $s^2 = 0.0455$. For the Gibbs sampler, we burn-in the algorithms for 10,000 iterations, and then save equally spaced samples of 5,000 draws from a longer run of 100,000. It takes only about one minute to run a two-factor model on my MacBook Pro with a 2.6GHz Intel Core i7 processor, 8 GB 1600 MHz DDR3 Memory running a Mac OS X Lion 10.7.5. The posterior means of Σ and β in a two-factor model are $E(\Sigma|y) = \text{diag}(0.05, 0.13, 0.62, 0.04, 0.25, 0.26)$ and

$$E(\beta'|y) = \begin{pmatrix} \mathbf{1.00} & \mathbf{0.96} & 0.46 & 0.39 & 0.42 & 0.41 \\ 0.00 & 0.05 & 0.43 & \mathbf{0.92} & \mathbf{0.78} & \mathbf{0.78} \end{pmatrix}.$$

Apart from the third row of $E(\beta|y)$, corresponding to the Japanese yen, which is equally explained by both common factors, one can argue that the first common factor groups North American currencies (US and Canadian dollars) and the second common factor groups European currencies. In factor analysis, it is fairly standard to summarize the importance of a common factor by its percentage contribution to the variability of a given attribute. Figure 5.2 presents the variance decomposition for the example and enhances the above statements regarding the interpretation of the two latent factors. Finally, Figure 5.3 illustrates the multimodality implied by overfitting the number of factors, regardless of using more or less informative prior specifications.

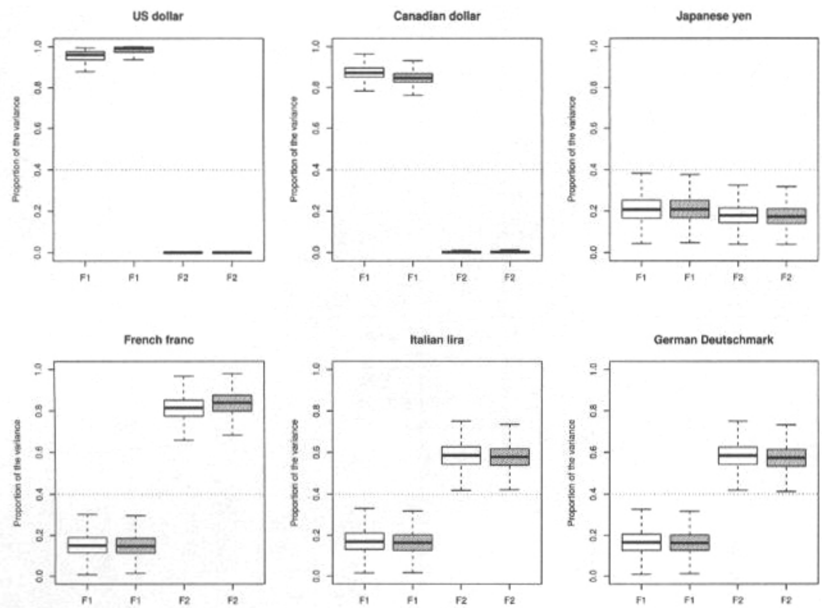

Figure 5.2: Example 1 – variance decomposition. The proportion of the variance of currency i, for $i = 1, \ldots, 6$, attributed to common factor j, for $j = 1, 2$ is given by $\nu_{ij} = \beta_{ij}^2/(\beta_{i1}^2 + \beta_{i2}^2 + \sigma_i^2)$. The boxplots summarize the posterior distributions of the ν_{ij}s. Informative prior (white boxplots). Noninformative prior (gray boxplots).

5.2.5 Number of Factors

Bayesian and non-Bayesian references that tackle the estimation/choice of the number of common factors are, amongst many others, Lawley and Maxwell (1963), Jöreskog (1967), Martin and McDonald (1975), Bartholomew (1981), Press (1982) (Chapter 10), Lee (1981), Akaike (1987), Bartholomew (1987), Press and Shigemasu (1989), Press and Shigemasu (1994). The book by Bartholomew (1987) is an excellent overview of the field right before MCMC tools became available.

Polasek (1997) uses Chib's methods (Chib, 1995) to approximate marginal likelihoods, and therefore posterior model probabilities, by running MCMC methods for factor models with a different number of common factors. Lopes and West (2004) introduced, developed and explored MCMC methods for factor models that treat the number of factors as unknown. Building on prior work on MCMC methods for a given number of factors, we introduced a reversible jump Markov chain Monte Carlo (RJMCMC, see Green, 1995)

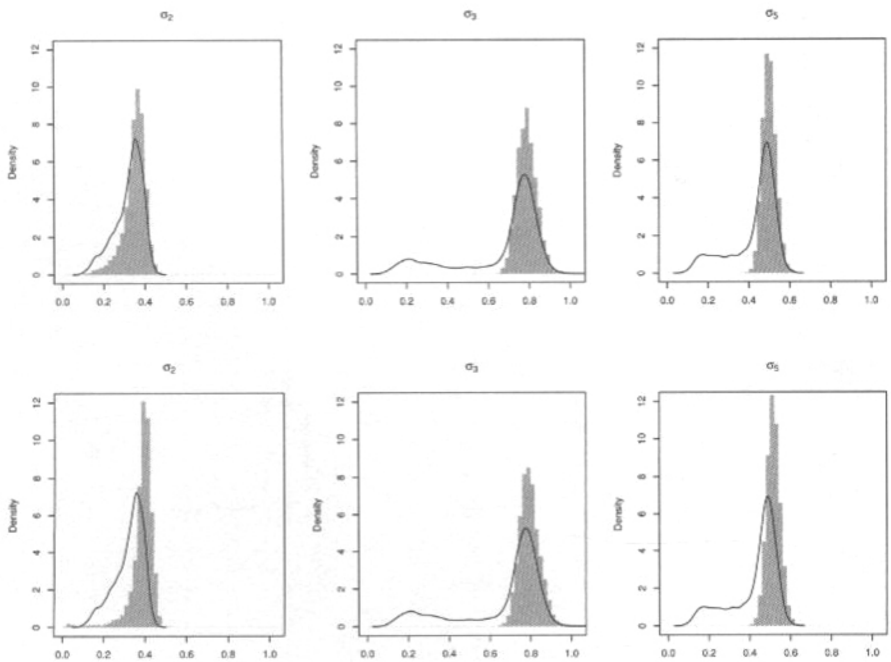

Figure 5.3: Example 1 – posterior inference. Posterior distributions (MCMC approximation) of $p(\sigma_i|y)$, $i = 2, \ldots, 5$ of Example 5.1, when estimating a two-factor model (histograms) and a (overfitted) three-factor model (solid lines). *Informative prior (top row):* The prior hyperparameters are $(m_0, C_0, \nu_0, s^2) = (0, 1, 2.2, 0.0455)$. Prior mode and median of σ_i are 0.154 and 0.252, respectively. *Noninformative prior (bottom row):* The prior hyperparameters are $(m_0, C_0, \nu_0, s^2) = (0, \infty, 0.001, 1)$. Prior mode and median of σ_i^2 are 0.032 and $1.4e^{149}$, respectively.

algorithm for moving between models with different numbers of factors, which avoids the computation of marginal likelihoods by treating the number of factors as a parameter.

Lopes and West (2004) served as the motivation to a handful of significant contributions to the discussion regarding the estimation of the number of common factors. West (2003) and Carvalho et al. (2008), for example, introduced high-dimensional factor analysis for modeling gene expression data with highly sparse factor loadings. The sparse representation induces a probability distribution over the number of factors. They model the factor scores via a Dirichlet. See also Lee and Song (2002) and Chow et al. (2011). Lopes et al.

(2008) extend Lopes and West's (2004) RJMCMC to the context of normal dynamic factor analysis, particularly when the loadings are spatially modeled. Bhattacharya and Dunson (2011), also focusing on high-dimensional problems, propose a gamma process shrinkage prior on the factor loadings matrix that allows the introduction of infinitely many factors of decaying importance.

We proposed, in Frühwirth-Schnatter and Lopes (2009), a new stochastic search strategy for estimating the number of factors and show that our approach encompasses and improves upon Carvalho et al. (2008), Bhattacharya and Dunson (2011) and most of the existing strategies. In Conti et al. (2011), we use and extend the new the strategy to examine the effect of early-life conditions and education on health but incorporating discrete attributes as well as limited dependent variable structures, commonly present when dealing with endogeneity in microeconometric studies.

5.3 Factor Stochastic Volatility

The basic, and certainly the most used and cited, *stochastic volatility* (SV) model can be described by the following non-linear dynamic model (West and Harrison, 1997):

$$
\begin{aligned}
y_t &= \exp\{x_t/2\}\varepsilon_t, && (5.7)\\
x_t &= \beta_0 + \beta_1 x_{t-1} + \tau\eta_t, && (5.8)
\end{aligned}
$$

where y_t are log-returns and log-variances $x_t = \log v_t$, ε_t and η_t i.i.d. standard normal errors. We take $\mu = 0$ for simplicity, $\beta_0 = \kappa\gamma$, $\beta_1 = 1 - \kappa$. The initial log-volatility state $x_0 \sim N(m_0, C_0)$, for known prior moments m_0 and C_0. An alternative specification assumes that $(x_0|\beta_0, \beta_1, \tau^2) \sim N(\beta_0/(1-\beta_1), \tau^2/(1-\beta_1^2))$ with $|\beta_1| < 1$; see Kalayloglu and Ghosh (2009) for Bayesian unit root tests regarding β. The centering parameterization moves β_0 to the observation equation and centers log-variances. This parameterization only marginally affects posterior inference in most cases while creating an unnecessary computational burden. We will then keep the simpler, less restrictive, more general specification with m_0 and C_0.

The SV model is completed with a conjugate prior distribution for $\theta = (\beta, \tau^2)$, i.e., $p(\theta) = p(\beta|\tau^2)p(\tau^2)$, where $(\beta|\tau^2) \sim N(b_0, \tau^2 B_0)$ and $\tau^2 \sim IG(c_0, d_0)$, for known hyperparameters b_0, B_0, c_0 and d_0. An alternative specification where β and τ^2 are independent *a priori* can be easily implemented with negligible additional computational cost.

Given a set of observed asset returns $y^n = (y_1, \ldots, y_n)$ and equations (5.7) and (5.8), the posterior distribution of the hidden volatility states and parameters (x^n, θ) is given by Bayes rule

$$
p(x^n, \theta|y^n) \propto p(\theta) \prod_{t=1}^{n} p(y_t|x_t, \theta)p(x_t|x_{t-1}, \theta), \tag{5.9}
$$

which is analytically intractable because of the nonlinearity of equation (5.7). Jacquier, Polson and Rossi (JPR, 1994) performed fully Bayesian inference through an MCMC scheme, with Kim et al. (1998) improving upon JPR's scheme based on the well known forward filtering-backward sampling (FFBS) algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994). Also, Jensen (2004) develops semiparametric inference for long-memory SV processes, while So et al. (1998) and Carvalho and Lopes (2004) accommodate Markov jumps in the log-volatilities.

5.3.1 Factor Stochastic Volatility

The literature on multivariate stochastic volatility models is now abundant, with Harvey et al. (1994), Pitt and Shephard (1999), Aguilar and West (2000), Lopes and Migon (2002), Chib et al. (2006) and Lopes and Carvalho (2007) representing only a few. Roughly speaking, they model the levels (or first differences) of a set of (financial) time-series by a standard normal factor model (Lopes and West, 2004) in which both the commonfactor variances and the specific (or idiosyncratic) time-series variances are modeled as univariate or multivariate (of low dimension) SV processes. The main practical and computational advantage of the factor stochastic volatility (FSV) model is its parsimony, where all the variances and covariances of a vector of time-series are modeled by a low dimensional stochastic volatility structure dictated by common factors. It is fairly common to find that, for large vectors of time series, the number of common factors is usually one or two orders of magnitude smaller, which speeds up computation and estimation considerably.

In this more general context, the simple normal linear factor model of equations (5.2) and (5.3) are replaced by

$$
\begin{aligned}
(y_t|f_t, \beta_t, \Sigma_t) &\sim N(\beta_t f_t; \Sigma_t), &(5.10)\\
(f_t|H_t) &\sim N(0; H_t), &(5.11)
\end{aligned}
$$

where $H_t = \text{diag}(h_{1t}, \ldots, h_{kt})$ contain the variances of the factors and $\Sigma_t = \text{diag}(\sigma_{1t}^2, \ldots, \sigma_{pt}^2)$. The main, nontrivial departure from the standard normal linear factor model lies in the time-varying structure of β_t, Σ_t and H_t. Log idiosyncratic variances, $\eta_{it} = \log \sigma_{it}^2$, are modeled by first-order autoregressions, AR(1):

$$
(\eta_{it}|\eta_{i,t-1}, \tilde{\alpha}_i, \rho_i, \tau_i^2) \sim N(\tilde{\alpha}_i + \rho_i \eta_{i,t-1}, \tau_i^2), \tag{5.12}
$$

for $i = 1, \ldots, p$. This is one of the simplest but certainly the most used specification in the literature (Jacquier et al., 1994). Similarly, the k-dimensional vector of factors' log variances, $\lambda_t = (\lambda_{1t}, \ldots, \lambda_{kt})'$, where $\lambda_{it} = \log h_{it}^2$, is modeled by a first-order vector autoregression, VAR(1), as

$$
(\lambda_t|\lambda_{t-1}, \alpha, \phi, U) \sim N(\alpha + \phi\lambda_{t-1}; U), \tag{5.13}
$$

with correlated innovations characterized by the non-diagonal matrix U (see Aguilar and West, 2000). When U is a diagonal matrix, the above multivariate model is reduced to k univariate conditionally independent autoregressive models (Pitt and Shephard, 1999). Both Pitt and Shephard (1999) and Aguilar and West (2000) consider $\beta_t = \beta$ for all t time periods. Lopes and Migon (2002) and Lopes and Carvalho (2007) extend the previous works by modeling the evolution of the unconstrained loadings, $\tilde{\beta}_t$, as

$$(\tilde{\beta}_t|\tilde{\beta}_{t-1}, \zeta, \Theta, W) \sim N(\zeta + \Theta\tilde{\beta}_{t-1}; W), \tag{5.14}$$

therefore allowing changes in covariances that are not exclusively associated to changes in the individual factor variances.

Philipov and Glickman (2006a,b) extend the above FSV model (with $\Sigma_t = \Sigma$) and model H_t as a full covariance matrix via their Wishart random process. They implement their model on return series 324 monthly observations of 88 individual companies from the S&P500 and use $k = 2$ common factors. Han (2006) implements a similar FSV model to form a portfolio based on 36 stocks, 1200 observations collected from the Center for Research in Security Prices (CRSP). Chib et al. (2006) introduce fat-tailed errors and jumps in the FSV model as well as efficient and fast MCMC algorithm. They implement their extension to simulated data ($p = 50$) and real data on international weekly stock index returns where $p = 10$ (see also Nardari and Scruggs, 2007). Lopes and Carvalho (2007) extend the FSV model to allow for Markovian regime shifts in the dynamic of the variance of the common factors and apply their model to study Latin America's main markets ($p = 5$). More recently, Nakajima and West (2013) extend the basic factor stochastic volatility model by allowing time-varying patterns of occurrence of zero elements in factor loadings matrices, which potentially leads to more interpretable, dynamic sparsity.

5.3.2 Financial Index Models

Carvalho et al. (2011) consider *financial index models* (FIM) appropriate choices for the purpose of covariance estimation and asset allocation. They develop a dynamic factor model encompass both the BARRA[1] and Fama-French[2] strategies in a simple yet flexible modeling setup. The fact that size, book-to-price and momentum are relevant to explain covariation among stocks is exploited in two common ways: (i) as individual regressors in a multivariate linear model, and (ii) as ranking variables used to construct portfolios that are used as indices. A very large body of literature is dedicated to selecting and testing the indices (Cochrane, 2001; Tsay, 2005).

[1]The BARRA strategy, after the company BARRA, Inc., founded by Barr Rosenberg, whose ideas can be found in Rosenberg and McKibben (1973).

[2]Fama and French, in a series of papers, identified a significant effect of market capitalization and book-to-price ratio into expected returns. This has lead to the now famous *Fama-French* three-factor model where, besides the market, two indices are built as portfolios selected on the basis of firms' size and book-to-price ratio.

FIM leads to tractable and parsimonious estimates of the covariances and it is economically interpretable and theoretically justified. From a methodological viewpoint, our models can be seen as a "structured" extension of current factor model ideas as developed in Aguilar and West (2000), West (2003), Lopes and West (2004), Lopes et al. (2008) and Carvalho et al. (2008). On the applied side our goal is to propose a model-based strategy that creates better FIM, helps deliver better estimates of time-varying covariances, and leads to more effective portfolios.

The general form of an index model assumes that stock returns follows $r_t = \alpha_t + \beta_t f_t + \varepsilon_t$, where, as in our general factor model of equation (5.2), f_t is a vector of common factors at time t, β_t is a matrix of factor loadings (or exposures) and ε_t is a vector of idiosyncratic residuals. Therefore, if $\text{Var}(f_t) = H_t$ and $\text{Var}(\varepsilon_t) = \Sigma_t$, then $\text{Var}(r_t) = \beta_t H_t \beta_t' + \Sigma_t$, from equation (5.4).

Carvalho et al. (2011) take a dynamic, model-based perspective and assume that at time t one observes the vector (r_t, x_t, Z_t), where r_t is a p-dimensional vector of stock returns; Z_t is a $p \times k$ matrix of firm specific information; and x_t is the market return (or some equivalent measure). The index model is then defined by a dynamic factor model as

$$r_t = \alpha_t + \gamma_t x_t + \beta_t f_t + \varepsilon_t, \tag{5.15}$$

where γ_t is a p-dimensional vector of market loadings, ε_t is the vector of idiosyncratic residuals, and f_t is a k-dimensional vector of common factors. Each element of both α_t and γ_t follows a standard first-order dynamic linear model (West and Harrison, 1997) and that ε_t is defined by a set of independent stochastic volatility models (Jacquier et al., 1994; Kim et al., 1998). They also assume that f_t is zero-mean multivariate normal with diagonal covariance matrix H_t driven by univariate stochastic volatility models. Finally, through β_t, company specific information will be used to help uncover relevant latent structures representing the risk factors. Taking $\gamma_t = 0$ and fixing the loading through time gets us to the factor stochastic volatility models of Section 5.3.

They consider five models, three of which we briefly list here for illustration. In all cases α_t and β_t follow standard first order dynamic linear models and ϵ_t follows standard stochastic volatility AR(1) models.

Dynamic CAPM: $\beta_t = 0$ and $r_t = \alpha_t + \gamma_t x_t + \varepsilon_t$.

Dynamic BARRA: β_t =(market size, book-to-price ratio, momentum) with $r_t = \alpha_t + \gamma_t x_t + \beta_t f_t + \varepsilon_t$ and $f_t \sim N(0, H_t)$.

Sparse Dynamic BARRA: We extend sparse factor analysis[3] to the dynamic BARRA by setting $\beta_{tj} = 0$ with probability $1 - \pi_{tj}$, where π_{tj}s are the inclusion probabilities.

[3]See West (2003), Carvalho et al. (2008) and Frühwirth-Schnatter and Lopes (2009).

5.4 Spatial Factor Analysis

In this section we discuss the extension of the basic normal linear factor model from equations (5.2) and (5.3) to accommodate spatially oriented data. Wang and Wall (2003), for instance, fit a spatial one-factor model to the mortality rates for three major diseases in nearly one hundred counties of Minnesota. In their case, y_i is the vector of p observed variables at each location s_i in region $\mathcal{D}$ and f_i is, as usual, the underlying common spatial factor at location s_i. The spatial structure comes in through the joint distribution of $f = (f_1, \ldots, f_n)'$ for the n locations:

$$f|\gamma \sim N(0, H(\gamma)), \tag{5.16}$$

where $H(\gamma)$ is the covariance matrix representing the spatial structure, and γ is the vector of parameters in the covariance structure. Two common covariance structures are the exponential model and the conditional autoregressive (CAR) model. In the exponential model, $H(\gamma)$ has components h_{ij} given by $h_{ij} = \alpha \exp\{-\phi d_{ij}\}$, where $d_{ij} = |s_i - s_j|$ is the distance between location s_i and s_j, α is the correlation-free variance and ϕ is a range parameter, which represents the decrease in correlation between two locations as the distance increases. Here, therefore, $\gamma = (\alpha, \phi)$. In the CAR model, f is discretely indexed over a partitioned area (areal data) and the correlations depend on the neighboring structure. In this case, $H(\gamma) = \tau^2(I_n - \rho W)^{-1}$, with spatial association parameter ρ, while τ^2 is the conditional variance of $f_i|f_{-i}$. W is a neighborhood matrix of the lattice with w_{ij} an indicator for whether areas i and j share a boundary. Here, therefore, $\gamma = (\rho, \tau^2)$. They also generalize their model to the Poisson common spatial factor analysis, which is applied to model the number of deaths due to lung, pancreas, esophagus and stomach cancers at the county level of Minnesota between the years of 1991 and 1998.

Christensen and Amemiya (2002, 2003) proposed what they called the shift-factor analysis method to model multivariate spatial data with temporal behavior modeled by autoregressive components, while Hogan and Tchernis (2004) and Lopes et al. (2012) fit a one-factor spatial model and entertained several forms of spatial dependence through the single common factor. See also Mezzetti and Billari (2005) and Mezzetti (2012) for other applications of Bayesian factor analysis to spatially correlated data in the contexts of socio-demographic and cancer incidence data.

The next two sections give further details on the spatially hierarchical factor model of Lopes et al. (2012), where we propose a model-based vulnerability index of the population from Uruguay to vector-borne diseases that combines different sources of information via a set of microenvironmental indicators and geographical location in the country. Lopes et al. (2008) propose a new class of nonseparable and nonstationary space-time models that resembles a standard dynamic factor model (Peña and Poncela, 2004, for instance), where the temporal dependence is modeled by latent factors while the spatial dependence is modeled by the factor loadings. Our model is applied to model the space-time structure of pollutants in the northern US.

5.4.1 Spatially Hierarchical Factor Analysis

We propose, in Lopes et al. (2012), a model-based vulnerability index of the population from Uruguay to vector-borne diseases. More specifically, the index is derived form a spatially oriented hierarchical factor model structure that combines both (Departamental) capital level data and census tract level data (see Figure 5.4). Apart form Bella Unión and Montevideo, with 11 and 1031 census tracts, respectively, the number of census tracts per capital varies roughly between 20 and 40. The $p = 11$ indicators used to construct the one-factor (vulnerability) index are standardized to represent percentages, averages, and densities observed at the census tract level: illiteracy rate, percentage of the population with access to public health care care, unemployed males, owed houses, households headed by a woman, households without sewage system, average number of persons per household, households with more than two persons per room, households without access to treated, drinkable water, households with air conditioner, and households poorly built.

For each one of the n_i census tracts of capital i, a p-dimensional vector of variables (social-economical, environmental, demographical, etc.) is observed, namely $y_{ij} = (y_{ij1}, \ldots, y_{ijp})'$, for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$. The spatially hierarchical factor model (SHFM) we proposed can be described according to the following hierarchy levels:

Observational level: Observations y_{ink} are linked to the vulnerability factor f_{ij},

$$y_{ijk} = \mu_k + \beta_k f_{ij} + \sigma_k \varepsilon_{ijk} \quad k = 1, \ldots, p, \tag{5.17}$$

where μ_k represents the overall grand mean vector for measurement k. The factor loadings vector $\beta = (1, \beta_2, \ldots, \beta_p)'$ plays an important role in understanding the role and the composition of the common factor. Its first element is set to one in order to ensure likelihood identifiability (see Lopes and West, 2004). The specific factors ε_{ijk} are standard normally distributed and independent across capitals, census tracts and measurements.

Vulnerability index level: The vector of factors $f_i = (f_{i1}, \ldots, f_{in_i})'$ within capital i is decomposed as the sum of two spatially structured components: one that captures the overall mean of the capital, and the other one captures the local structure of the index, in the census tracts' level, and also accounts for possible effects of neighboring census tracts. More precisely, we assume

$$f_{ij} = \theta_i + \tilde{f}_{ij} + \sqrt{\omega_i} u_{ij}, \tag{5.18}$$

Figure 5.4: Spatially hierarchical factor analysis. The state level data is summarized by the 19 departmental capitals. Census tract data is exemplified here by the city of Melo, which is the capital of Cerro Lago. This figure reproduces Figure 1 from Lopes et al. (2012).

where θ_i is the common factor for capital i, $\tilde{f}_{ij}$ is the specific factor for census tract j and capital i, and u_{ij} are independent standard normals. The error term u_{ij} accounts for unanticipated, location specific idiosyncrasies.

Within capital variation: As the capitals are divided into census tracts defining irregular subregions, we model the within capital factors $\tilde{f}_i = (\tilde{f}_{i1}, \ldots, \tilde{f}_{in_i})'$, for $i = 1, \ldots, I$, by a proper conditionally autoregressive (CAR) specification (Sun, Tsutakawa, and Speckman, 1999):

$$\tilde{f}_i \sim N(0, \tau_i^2 P_i), \tag{5.19}$$

where $P_i = P_i(\phi) = (I_{n_i} + \phi M_i)^{-1}$, $M_i = D_i - W_i$, with w_{ijl}, the (j, l) component of W_i, given by $w_{ijl} = 1/d_{jl}$ if j and l are neighbors (denoted here by $j \sim l$) and zero otherwise, $d_{jl} = ||s_j - s_l||$ is the Euclidean distance between centroids of capitals j and l, $D_i = \text{diag}(w_{i1+}, \ldots, w_{in_i+})$ and $w_{ij+} = \sum_{l \sim j} w_{ijl}$. The inverse matrix P_i^{-1} is diagonally dominant and positive definite (Harville, 1997). The parameter ϕ controls the strength of the association between the components of $\tilde{f}_i$, with $\phi = 0$ implying independence. Equation (5.19) approaches the intrinsic autoregressive model when ϕ approaches infinity (Besag, York, and Mollié, 1991; Besag and Kooperberg, 1995).

Between capitals variation: The θ_is are conditionally independent and Gaussian with common baseline vulnerability factor θ_0 and covariance structure driven by the Euclidean distances between the centroids of the capitals, i.e.,

$$\theta \sim N\left(1_I\theta_0, \delta^2 H(\lambda)\right), \tag{5.20}$$

where $\theta = (\theta_1, \cdots, \theta_I)$. Although each capital i has its own vulnerability factor, the above model allows borrowing-strength across neighboring regions. The correlation matrix H is fully specified by a Matérn correlation function, i.e., $H_{ij} = \rho(\lambda, d_{ij}) = 2^{1-\lambda_2}\Gamma(\lambda_2)^{-1}(d_{ij}/\lambda_1)^{\lambda_2}\mathcal{K}_{\lambda_2}(d_{ij}/\lambda_1)$ where $\mathcal{K}_{\lambda_2}$ is the modified Bessel function of the second kind and of order λ_2, $\lambda = (\lambda_1, \lambda_2)$ and $d_{ij} = ||s_i - s_j||$ is the Euclidean distance between the centroids s_i and s_j of capitals i and j, for $i, j = 1, \ldots, I$.

Figure 5.5 illustrates the performance of the SHFM in producing the vulnerability indexes to all 1031 census tracts of Montevideo. The figure helps discriminating very opposing situations within the city and captures the local effects of the factor/index. In other words, the richest Southeast area of the city has fewer census tracts with low values of the index, representing low vulnerability. Land in these regions is irregularly owned. Overall, the levels of the index in Montevideo are in accordance to what is anticipated by experts, showing high values (more vulnerability) towards the North-West region which comprise more rural areas.

The next model uses factor analysis to reduce the dimension of multiple time-series representing the dynamics of dozens of locations. In this case, the attributes are, in fact, univariate measurements at the locations. We will see that the structure of the matrix of factor loadings plays an important role in capturing conditional spatial variation.

5.4.2 Spatial Dynamic Factor Analysis

Lopes et al. (2008) propose a new class of nonseparable and nonstationary space-time models that resembles a standard dynamic factor model (Peña and Poncela, 2004, for instance),

$$\begin{aligned}
y_t &= \mu_t^{y*} + \beta f_t + \epsilon_t, & \epsilon_t &\sim N(0, \Sigma), & (5.21)\\
f_t &= \Gamma f_{t-1} + \omega_t, & \omega_t &\sim N(0, \Lambda), & (5.22)\\
\beta_{(j)} &\sim GRF(\mu_j^{\beta*}, \tau_j^2 \rho_{\phi_j}(\cdot)) \equiv N(\mu_j^{\beta*}, \tau_j^2 R_{\phi_j}), & & & (5.23)
\end{aligned}$$

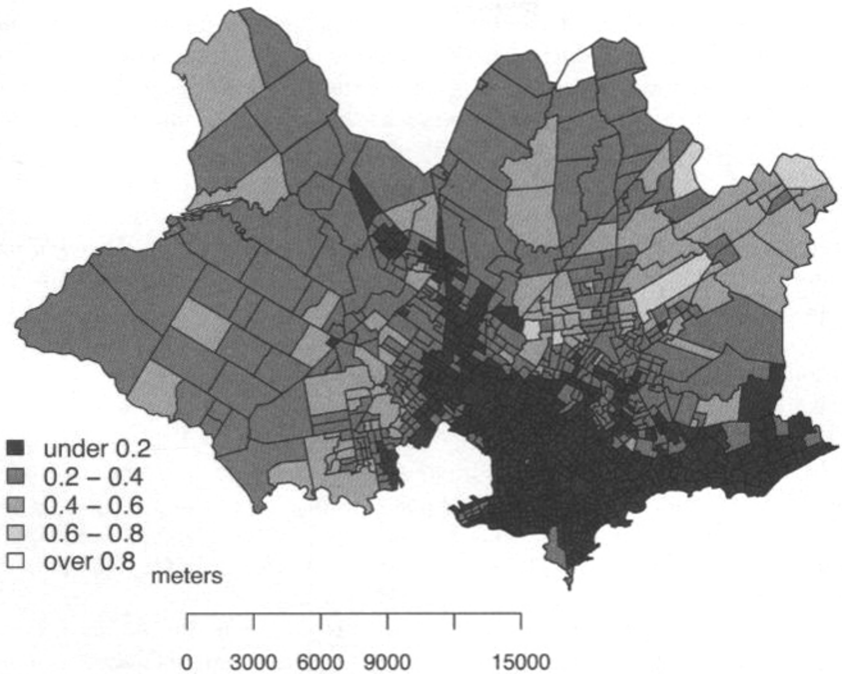

Figure 5.5: Spatially hierarchical factor analysis. Within-city posterior (standardized) vulnerability index per census tract of Montevideo. This is Figure 5 from Lopes et al. (2012).

where $y_t = (y_{1t}, \ldots, y_{Nt})'$ is the N-dimensional vector of observations (locations $s_1, \ldots, s_N$ and times $t = 1, \ldots, T$), μ_t^{y*} is the mean level of the space-time process, f_t is an m-dimensional vector of common factors, for $m < N$ (m is potentially several orders of magnitude smaller than N) and the jth column of the factor loadings matrix,

$$\beta_{(j)} = (\beta_{(j)}(s_1), \ldots, \beta_{(j)}(s_N))',$$

for $j = 1, \ldots, m$, is modeled as a conditionally independent, distance-based Gaussian process or a Gaussian random field (GRF). The matrix Γ characterizes the evolution dynamics of the common factors, while Σ and Λ are observational and evolutional variances. $\mu_j^{\beta*}$ is a N-dimensional mean vector. The (l, k)-element of R_{ϕ_j} is given by $r_{lk} = \rho_{\phi_j}(|s_l - s_k|)$, $l, k = 1, \ldots, N$, for suitably defined correlation functions $\rho_{\phi_j}(\cdot)$, $j = 1, \ldots, m$, for instance, exponential, power exponential or Matérn. The parameters ϕ_js are typically scalars or low dimensional vectors (for details, see Cressie, 1993 and Stein, 1999).

The univariate measurements from all observed locations, either areal or point-referenced, at any given time, are grouped together in what otherwise would be the vector of attributes in standard factor analysis and the spatial dependence is introduced by the columns of the factor loadings matrix. Therefore, common dynamic factors can be thought of as describing temporal similarities amongst the time series, such as common annual cycles or (stationary or nonstationary) trends, while the importance of common factors in describing the measurements in a given location is captured by the components of the factor loadings matrix. More general time series models can be entertained, through the common factors, without imposing additional constraints to the current spatial characterization of the model, and vice-versa (Lopes et al., 2008).

Lopes et al. (2008) model the spatial and temporal variations in the concentration levels of sulfur dioxide, SO_2, across 24 monitoring stations. Weekly measurements in $\mu g/m^3$ are collected by the Clean Air Status and Trends Network (CASTNet), which is part of the Environmental Protection Agency (EPA) of the United States. Measurements span from the first week of 1998 to the 30th week of 2004, a total of 342 observations. Figure 5.6 shows the posterior mean surface interpolation corresponding to the seasonal factor when fitting one of their SDFM. The loadings for the seasonal factor are smaller in the highly industrialized state of Ohio.

Calder (2007) also used SO_2 along with three other pollutants in a related dynamic factor model where the columns of the factor loadings matrix is model via deterministic smoothed kernels (see also Sansó et al., 2004). Several other models are particular cases of the our SDFM: principal kriging (Sahu and Mardia, 2005, and Lasinio, Sahu and Mardia, 2005), kriged Kalman filter (Mardia et al., 1998), and orthonormal basis loadings (Wikle and Cressie, 1999).

5.5 Additional Developments

5.5.1 Prior and Posterior Robustness

Lee and Press (1998) studies posterior robustness of the loadings, common factors and idiosyncratic covariance, while Lopes (2003) studies prior specification and sensitivity in factor models via the expected posterior prior setup of Pérez and Berger (2002).

5.5.2 Mixture of Factor Analyzers

Mixture of factor analyzers (MFA) is a nonlinear, more flexible extension of the linear factor analysis of Section 5.2. The basic structure of a MFA model is given by

$$(y|\beta, f, \Sigma) \sim \sum_{j=1}^{m} \pi_j N(y; \beta_j f; \Sigma), \tag{5.24}$$

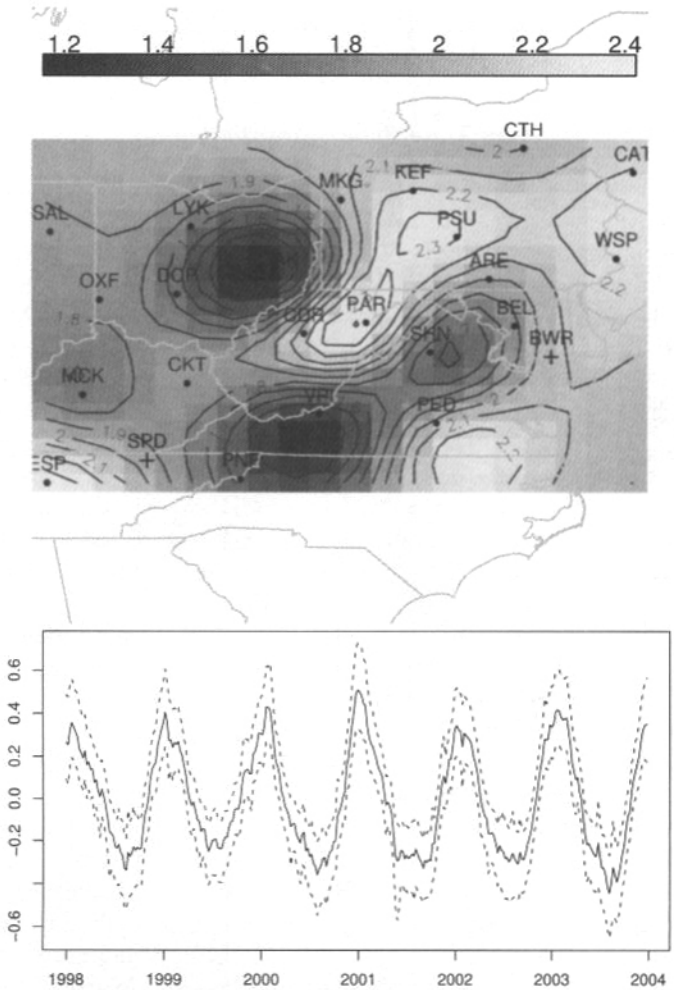

Figure 5.6: Spatial dynamic factor model. Posterior Bayesian interpolation for loadings factors or the northeast part of the U.S. (top frame). Posterior means (and 95% credibility intervals) of the seasonal common factors (bottom frame). These are taken from Figures 5 and 6 from Lopes et al. (2008).

where m is the number of *analyzers*. Conditional on j, a standard normal linear factor model arises. Ghahramani and Beal (2000) introduce an algorithm that fits mixture of factor analyzers models via variational approximation to full Bayesian integration over model parameters. Utsugi and Kumagai (2001), Fokoué and Titterington (2003) and Fokoué (2004) propose MCMC-based posterior inference for MFA models. McLachlan et al. (2007) and Andrews and

McNicholas (2011) extend the MFA to the multivariate t family and uses the Expectation-Maximation (EM) algorithm for parameter estimation.

5.5.3 Factor Analysis in Time Series Modeling

Peña and Box (1987) introduce factor models with common, independent or dependent, factors following ARMA processes. Similarly, Engle (1987) proposes a multivariate ARCH with factor structures. Diebold and Nerlove (1989) model multivariate GARCH structures through a one-factor model to study the dynamics of exchange rate volatilities. Engle et al. (1990), who use Factor-ARCH to model a conditional covariance matrix of asset returns, while Ng and Rothschild (1992) relates dynamic and static factors to portfolio allocation in financial markets. Lin (1992) compares four frequentist estimators for factor GARCH models: two-stage univariate GARCH, two-stage quasi-maximum likelihood, quasi-maximum likelihood with known factor weights and quasi-maximum likelihood with unknown factor weights. Molenaar et al. (1992) employ nonstationary multivariate time series dynamic factor analysis with lagged common factors to account for the persistence in time series trends.

Bollerslev and Engle (1993) introduce a k-factor GARCH model and study co-persistence in multivariate integrated GARCH models. Harvey et al. (1994) introduce one of the first factor stochastic volatility models (see Section 5.3.1). Escribano and Peña (1994) establish the relationship between cointegrated vectors and common factors via Peña and Box's (1987) dynamic factor models. Demos and Sentana (1998) introduces an EM algorithm for conditionally heterescedastic factor models, while Sentana (1998) investigates the similarities and the differences of Engle's (1987) factor GARCH model and Diebold and Nerlove's (1989) latent factor ARCH model. See also Fiorentini et al. (2004). Vrontros et al. (2003) proposes a full-factor multivariate GARCH model (2003) and treats the order of variable via Bayesian model averaging. More recently, Sentana et al. (2008) derive indirect estimators of conditionally heteroskedastic factor models in which the volatilities of common and idiosyncratic factors depend on their past unobserved values. See also the recent paper by Zhou et al. (2012) that introduces correlated dynamic latent factor structures into a new class of latent threshold dynamic factor models for multivariate volatility analysis and forecasting of financial time series.

5.5.4 Factor Analysis in Macroeconometrics

Stock and Watson (2002b,2002a) implement principal components analysis project a large number of predictors (about 215 in their case) on a few principal components, or *diffusion indexes*. Then, the diffusion indexes are used as explanatory variables when forecasting a macroeconomic time series variable. More precisely, let z_{t+1} donate the macroeconomic time series and y_t a p-dimensional vector with (possibly many highly correlated) predictors. They assume that (z_{t+h}, y_t) admit a factor model representation with k common factors f_t, whose simplest version is

$$y_t = \beta f_t + \varepsilon_t, \tag{5.25}$$
$$z_{t+h} = \theta' f_t + u_{t+h}. \tag{5.26}$$

Roughly speaking, they propose a two-step estimation procedure. First, the diffusion indexes are estimated from Equation (5.25) by principal component analysis. Second, the estimated diffusion indexes are plugged in the forecasting Equation (5.26). Their empirical applications aim at forecasting major monthly macroeconomic variables for the United States (1959:1 to 1998:12), such as total industrial production, real personal income less transfers, real manufacturing and trade sales, and number of employees on nonagricultural payrolls. The predictors represent main categories of macroeconomic time series, including real output and income, employment and hours, real retail, manufacturing, and trade sales and consumption, amongst many others. They showed empirically that only $k = 6$ factors account for much of the variance of our $p = 215$ time series. Bernanke et al. (2005) and Stock and Watson (2005) combined factor models with vector autoregressive models, commonly known as *factor-augmented VAR* (FAVAR) models, while Negro and Otrok (2008) and Korobilis (2013) extend these models, from a Bayesian viewpoint, to include time-varying coefficients. The review paper by Stock and Watson (2006) provides an important review to the area of forecasting with many predictors.

The Bayesian approach to factor analysis applied to macroeconomic problems has grown considerably over the last decade. A few important contributions are the following. Koop and Potter (2004) revisits Stock and Watson (2002b) and implement Bayesian model averaging on the above dynamic factor structure. Otrok and Whiteman (1998) designs and implements a Bayesian dynamic latent factor and produce coincident and leading indicators for a local US economy based on the posterior mean of the latent factors. Kose et al. (2003) proposes a dynamic factor model for international business cycles whose common factors are divided into world, region and country specific ones (see also Koop and Korobilis, 2009). Uhlig and Ahmadi (2012) proposes a Bayesian factor-augmented VAR, or BFAVAR, to study the effects of monetary policy shocks. A recent overview of Bayesian macroeconometrics is provided by Del Negro and Schorfheide (2011).

5.5.5 Term Structure Models

The yield curve is defined as the relationship between τ and $\tau^{-1} \log p_t(\tau)$, where $p_t(\tau)$ is the price, at time t, of a zero-coupon bond with payoff 1 at maturity $t + \tau$. See, for instance, Diebold and Li (2006) and Diebold et al. (2008) for further details. Diebold and Li (2006), for example, argue that two popular approaches to the term structure modeling are *(i)* no-arbitrage models and *(ii)* equilibrium models, with significant contributions to the former by Hull and White (1990) and Heath et al. (1992), and to the latter by Vasicek (1977), Cox et al. (1985) and Duffie and Kan (1996), amongst others.

Diebold and Li (2006), however, use variations of the Nelson and Siegel (1987) exponential components framework to model the yield curve as a three-factor model (level, slope and curvature) that evolves over time dynamically. Nelson-Siegel framework imposes a parsimonious structure on the factor loadings matrix:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right), \tag{5.27}$$

with λ_t characterizing the decaying rate; small values of λ_t produce slow decay and can better fit the curve at long maturities, while large values of λ_t produce fast decay and can better fit the curve at short maturities.

Chib and Ergashev (2009) presented a Bayesian approach for the fitting of affine yield curve models with macroeconomic factors that emphasizes the use of a prior on the parameters of the model which implies an upward-sloping yield curve. Chib and Kang (2012) extend the model to accommodate change-points in affine term structure models.

5.5.6 Sparse Factor Structures

We finish the chapter by summarizing the current literature on Bayesian sparse factor analysis. We have already listed a few of these contributions, such as West (2003), Carvalho et al. (2008) and Frühwirth-Schnatter and Lopes (2009). Mayrink and Lucas (2013), for instance, study gene expression data by extending Carvalho et al. (2008) to entertain interactions between the common factors in their sparse factor model. Similarly, Conti et al. (2011) implement Frühwirth-Schnatter and Lopes' (2009) parsimonious factor analysis strategy to examine the effect of early-life conditions and education on health but incorporating discrete attributes as well as limited dependent variable structures, commonly present when dealing with endogeneity in microeconometric studies.

Additional references are Hahn et al. (2012), Pati et al. (2012), Cron and West (2012) and Hahn et al. (2013), who propose sparse factor probit modeling, sparse factor analysis for massive covariance matrices, random sparse orthogonal matrices, and predictor-dependent shrinkage partial factor analysis, respectively. See also Yoshida and West (2013) for sparse graphical factor models, and Knowles and Ghahramani (2011) for nonparametric Bayesian sparse factor models applied, once again, to gene expression modeling. Sparse dynamic factor models are proposed, for instance, by Kaufmann and Schumacher (2013) who introduce a new, order-independent identification strategy based on semi-orthogonal loadings. See also Zhang and Nesselroade (2007), Kaufmann and Schumacher (2012) and Cui and Dunson (2012).

5.6 Modern non-Bayesian factor analysis

The literature on modern factor analysis is growing, as Table 5.1 suggested, as expected outside the realm of Bayes. For those more eager to explore more non-Bayesian factor analysis alternatives, the following paragraphs bring some of the recent papers by Bai and Ng's group and by Forni, Hallin, Lippi and Reichlin's groups. The list is narrow and limited, which reflects the author's own limitations. These few papers, as well as their reference lists, we believe, will provide the reader with a start-up for her own search.

Bai (2003), Bai and Li (2012) and Li (2013) consider ML estimation of high-dimensional (static and dynamic) factor models when p is at least as large as n. Bai and Ng (2002) propose tools to select the number of common factor in the above large n and large p scenario. Bai and Ng (2008) survey the main theoretical results, including how to determine the number of factors, how to conduct inference factor-regression models (see also Bai and Ng, 2006, Amengual and Watson, 2007). Bai and Ng (2010) propose a class of factor instrumental variables in the context of data rich environment, where a large number of endogenous variables are driven by a small number of unobservable exogenous common factors. Moench et al. (2011) propose a (inherently Bayesian) large dimensional hierarchical factor model that takes into account within- and between-block variability. See also Fan et al. (2008), Fan et al. (2011), Onatski (2009) and Onatski (2012).

Forni et al. (2000), Forni and Lippi (2001), Forni et al. (2005), Forni and Lippi (2011) introduce generalized (dynamic) factor models and discuss extensively identifiability, estimation and forecasting, while Forni et al. (2009) talk structural factor model with large cross-sections. Hallin and Liska (2007) tackle the estimation of the number of common factors in the above generalized dynamic factor model, while Hallin and Liska (2011) decompose large sets of macroeconomic variables into smaller, but still quite large, blocks. Additional works tackling high dimensional (dynamic) factor models in econometrics are Kapetanios and Marcellino (2009), Doz et al. (2011, 2012) and Hallin et al. (2011).

5.7 Final Remarks

We started reviewing the basic normal linear factor model, which is considered the spinal cord of all factor models presented later on in the chapter. The one hundred plus references listed at the end of the chapter represent a biased fraction of the existing literature. Factor models appear now in virtually all areas of sciences, despite its start in psychometric studies back in the 1900s with the seminal work of Spearman. Amongst its various extensions, we discussed and referenced several factor stochastic volatility models, dynamic factor models, spatial factor models and sparse factor models.

We believe the next several years will witness the multiplication of more complex and massively large factor structures and applications to study related objects other than neatly organized rectangle matrices of observations (rows) and attributes (columns). The interactions between statisticians, econometricians, machine learners and data miners will lead to highly sophisticated and computationally efficient and fast factor-based modeling.

Bibliography

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics 18*, 338–357.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika 52*, 317–332.

Amemiya, Y. and T. W. Anderson (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics 18*, 1453–1463.

Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large n and t panel. *Journal of Business and Economic Statistics 25*, 91–96.

Anderson, T. W. and Y. Amemiya (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics 16*, 759–771.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability*, Volume 5, pp. 111–150. University of California Press, Berkeley.

Andrews, J. L. and P. D. McNicholas (2011). Extending mixtures of multivariate t-factor analyzers. *Statistics and Computing 21*, 361–373.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–71.

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics 40*, 436–465.

Bai, J. and S. Ng (2002). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–171.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*, 1133–1150.

Bai, J. and S. Ng (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics 3*, 89–163.

Bai, J. and S. Ng (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory 26*, 1577–1606.

Bartholomew (1995). Spearman and the origin and development of factor analysis. *Journal of Mathematical and Statistical Psychology 48*, 211–220.

Bartholomew, D. (1981). Posterior analysis of the factor model. *The British Journal of Mathematical and Statistical Psychology 34*, 93–99.

Bartholomew, D. (1987). *Latent Variable Models and Factor Analysis*. London: Charles Griffin.

Bentler, P. M. and J. S. Tanaka (1983). Problems with EM algorithms for ML factor analysis. *Psychometrika 48*, 247–251.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics 7*, 387–422.

Besag, J. and C. Kooperberg (1995). On conditional and intrinsic autoregressions. *Biometrika 82*, 733–746.

Besag, J., J. C. York, and A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics 43*, 1–59.

Bhattacharya, A. and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika 98*, 291–306.

Bollerslev, T. and R. F. Engle (1993). Common persistence in conditional variances. *Econometrica 61*, 167–186.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley.

Broemeling, L. (1985). *Bayesian Analysis of Linear Models*. New York: Marcel Dekker.

Burt, C. (1940). *The Factors of the Mind: An Introduction to Factor Analysis in Psychology*. University of London Press, London.

Calder, C. (2007). Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. *Environmental and Ecological Statistics 14*, 229–47.

Carter, C. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–553.

Carvalho, C. and H. Lopes (2004). Simulation-based sequential analysis of Markov switching stochastic volatility models. Technical report, Graduate School of Business, University of Chicago.

Carvalho, C. M., J. Chang, J. Lucas, Q. Wang, J. Nevins, and M. West (2008). High-dimensional sparse factor modelling: Applications in gene expression genomics. *Journal of the American Statistical Association 103*, 1438–1456.

Carvalho, C. M., H. F. Lopes, and O. Aguilar (2011). Dynamic stock selection strategies: A structured factor model framework. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 9*, Volume 9, pp. 69–90. Oxford University Press, Oxford.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. and B. Ergashev (2009). Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association 104*, 1324–1337.

Chib, S. and K. H. Kang (2012). Change-points in affine arbitrage-free term structure models. *Journal of Financial Econometrics (available online)*.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*, 341–371.

Chow, S.-M., N. Tang, Y. Yuan, X. Song, and H. Zhu (2011). Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior. *British Journal of Mathematical and Statistical Psychology 64*, 69–106.

Christensen, W. and Y. Amemiya (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association 97*(457), 302–317.

Christensen, W. and Y. Amemiya (2003). Modeling and prediction for multivariate spatial factor analysis. *Journal of Statistical Planning and Inference 115*, 543–564.

Cochrane, J. (2001). *Asset Pricing.* Princeton University Press.

Conti, G., J. J. Heckman, H. F. Lopes, and R. Piatek (2011). Constructing economically justified aggregates: an application of the early origins of health. Discussion paper, University of Chicago Booth School of Business.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

Cressie, N. (1993). *Statistics for Spatial Data.* New York: Wiley.

Cron, A. and M. West (2012). Models of random sparse eigenmatrices with application to Bayesian factor analysis. Discussion Paper 2012-01, Department of Statistical Science, Duke University.

Cudeck, R. and R. C. MacCallum (2007). *Factor Analysis at 100: Historical Developments and Future Directions.* Routledge.

Cui, K. and D. B. Dunson (2012). Generalized dynamic factor models for mixed-measurement time series. *Journal of Computational and Graphical Statistics*.

Del Negro, M. and F. Schorfheide (2011). Bayesian macroeconometrics. In J. Geweke, G. Koop, and H. V. Dijk (Eds.), *Handbook of Bayesian Econometrics.* Oxford University Press.

Demos, A. and E. Sentana (1998). An EM algorithm for conditionall heterescedastic factor models. *Journal of Business and Economic Statistics 16*, 357–361.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Diebold, F. X., C. Li, and V. Z. Yue (2008). Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics 146*, 3511–363.

Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent ARCH model. *Journal of Applied Econometrics 4*, 1–21.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics 164*, 188–205.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics 94*, 1014–1024.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*, 379–406.

Engle, R. (1987). Multivariate ARCH with factor structures - cointegration in variance. Discussion paper, University of California, San Diego, Department of Economics.

Engle, R. F., V. K. Ng, and M. Rothschild (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics 45*, 213–238.

Escribano, A. and D. Peña (1994). Cointegration and common factors. *Journal of Time Series Analysis 15*, 577–586.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*, 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics 39*, 3320–3356.

Fiorentini, G., E. Sentana, and N. Shephard (2004). Likelihood-based estimation of latent generalized ARCH structures. *Econometrica 72*, 1481–1517.

Fokoué (2004). Mixtures of factor analysers: An extension with covariates. *Journal of Multivariate Analysis 95*, 370–384.

Fokoué, E. and D. M. Titterington (2003). Mixtures of factor analysers: Bayesian estimation and inference by stochastic simulation. *Machine Learning 50*, 73–94.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models with large cross-sections. *Econometric Theory 25*, 1319–1347.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized factor model: Identification and estimation. *The Review of Economics and Statistics 82*, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model one-sided estimation and forecasting. *Journal of the American Statistical Association 100*, 830–840.

Forni, M. and M. Lippi (2001). The generalized factor model: Representation theory. *Econometric Theory 17*, 1113–1141.

Forni, M. and M. Lippi (2011). The general dynamic factor model: One-sided representation results. *Journal of Econometrics 163*, 23–28.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183–202.

Frühwirth-Schnatter, S. and H. F. Lopes (2009). Parsimonious Bayesian factor analysis when number of factors is unknown. Discussion paper, University of Chicago Booth School of Business.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference* (2nd ed.). Chapman & Hall/CRC.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies 9*, 557–587.

Ghahramani, Z. and M. Beal (2000). Variational inference for Bayesian mixtures of factor analysers. In S. A. Solla, T. K. Leen, and K. Müller (Eds.), *Advances in Neural Information Processing Systems*, Volume 12, pp. 449–455. MIT Press.

Ghosh, J. and D. Dunson (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics 18*, 306–320.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Hahn, P. R., C. M. Carvalho, and S. Mukherjee (2013). Partial factor modeling: Predictor-dependent shrinkage for linear regression. *Journal of the American Statistical Association 108*, 999–1008.

Hahn, P. R., C. M. Carvalho, and J. Scott (2012). A sparse factor analytic probit model for congressional voting patterns. *Applied Statistics 61*, 619–635.

Hallin, M. and R. Liska (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*, 603–617.

Hallin, M. and R. Liska (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics 163*, 29–41.

Hallin, M., C. Mathias, H. Pirotte, and D. Veredas (2011). Market liquidity as dynamic factors. *Journal of Econometrics 163*, 42–50.

Han, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies 19*, 237–271.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–264.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag, New York.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. *Econometrica 60*, 77–105.

Hogan, J. W. and R. Tchernis (2004). Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association 99*, 314–324.

Holzinger, K. J. and H. H. Harman (1941). *Factor analysis: A Synthesis of Factorial Methods.* University of Chicago Press, Chicago.

Hotelling, H. (1955). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology 24*, 417–441.

Hull, J. and A. White (1990). Pricing interest-rate-derivative securities. *Review of Financial Studies 3*, 573–592.

Ihara, M. and Y. Kano (1995). Identifiability of full, marginal, and conditional factor analysis model. *Statistics and Probability Letters 23*, 343–350.

Jacquier, E., N. Polson, and P. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics 12*, 371–389.

Jensen, M. (2004). Semiparametric Bayesian inference of long-memory stochastic volatility models. *Journal of Time Series Analysis 25*, 895–922.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika 32*, 443–382.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika 34*, 183–220.

Kapetanios, G. and M. Marcellino (2009). A parametric estimation method for dynamic factor models of large dimensions. *Journal of Time Series Analysis 30*, 208–238.

Kaufmann, S. and C. Schumacher (2012). Finding relevant variables in sparse Bayesian factor models: Economic applications and simulation results. Discussion Papers 29/2012, Deutsche Bundesbank, Research Centre.

Kaufmann, S. and C. Schumacher (2013). Bayesian estimation of sparse dynamic factor models with order-independent identification. Working Papers 13.04, Swiss National Bank, Study Center Gerzensee.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Knowles, D. and Z. Ghahramani (2011). Nonparametric Bayesian sparse factor models with application to gene expression modeling. *Annals of Applied Statistics 5*, 1534–1552.

Koop, G. and D. Korobilis (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*, 267–358.

Koop, G. and S. Potter (2004). Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal 7*, 550–565.

Korobilis, D. (2013). Assessing the transmission of monetary policy shocks using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics 75*, 157–179.

Kose, A., C. Otrok, and C. H. Whiteman (2003). International business cycles: World, region, and country specific factors. *American Economic Review 93*, 1216–1239.

Lasinio, G. J., S. K. Sahu, and K. V. Mardia (2005). Modeling rainfall data using a Bayesian kriged-Kalman model. Technical report, School of Mathematics, University of Southampton.

Lawley (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh 60*, 64–82.

Lawley (1953). Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh 60*, 64–82.

Lawley, D. and A. Maxwell (1963). *Factor Analysis as a Statistical Method.* London: Butterworth.

Lee, S. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika 46*, 153–160.

Lee, S. E. and S. J. Press (1998). Robustness of Bayesian factor analysis estimates. *Communications in Statistics, Theory and Methods 27*, 1871–1893.

Lee, S. Y. and X. Y. Song (2002). Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika 29*, 23–40.

Li, K. (2013). Fixed-effects dynamic panel models, a factor analytical method. *Econometrica 81*, 285–314.

Lin, W.-L. (1992). Alternative estimators for factor GARCH models - a Monte Carlo comparison. *Journal of Applied Econometrics 7*, 259–279.

Lopes, H. F. (2003). Expected posterior priors in factor analysis. *Brazilian Journal of Probability and Statistics 17*, 91–105.

Lopes, H. F. and C. M. Carvalho (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference 137*, 3082–3091.

Lopes, H. F. and H. S. Migon (2002). Comovements and contagion in emergent markets: Stock indexes volatilities. In C. Gatsonis, R. Kass, A. Carriquiry, A. Gelman, Verdinelli, D. Pauler, and D. Higdon (Eds.), *Case Studies in Bayesian Statistics, Volume 6*, pp. 287–302.

Lopes, H. F., E. Salazar, and D. Gamerman (2008). Spatial dynamic factor models. *Bayesian Analysis 5*, 1–30.

Lopes, H. F., A. Schmidt, E. Salazar, M. Gomez, and M. Achkar (2012). Measuring vulnerability via spatially hierarchical factor models. *Annals of Applied Statistics 6*, 284–404.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*, 41–67.

Mardia, K. V., C. R. Goodall, E. Redfern, and F. J. Alonso (1998). The kriged Kalman filter (with discussion). *Test 7*, 217–85.

Martin, J. and R. McDonald (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika 40*, 505–517.

Mayrink, V. D. and J. E. Lucas (2013). Sparse latent factor models with interactions: Analysis of gene expression data. *The Annals of Applied Statistics 7*, 799–822.

McLachlan, G., R. Bean, and L. Jones (2007). Extension of the mixture of factor analyzers model to incorporate the multivariate t-distribution. *Computational Statistics and Data Analysis 11*, 5327–5338.

Mezzetti, M. (2012). Bayesian factor analysis for spatially correlated data: application to cancer incidence data in Scotland. *Statistical Methods & Applications 21*, 49–74.

Mezzetti, M. and F. C. Billari (2005). Bayesian correlated factor analysis of socio-demographic indicators. *Statistical Methods & Applications 14*, 223–241.

Moench, E., S. Ng, and S. Potter (2011). Dynamic hierarchical factor models. *Review of Economics and Statistics (to appear).*

Molenaar, P. C. M., J. G. D. Gooijer, and B. Schmitz (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika 57*, 333–349.

Nakajima, J. and M. West (2013). Dynamic factor volatility modeling: A Bayesian latent threshold approach. *Journal of Financial Econometrics 11*, 116–153.

Nardari, F. and J. T. Scruggs (2007). Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and APT pricing restrictions. *Journal of Financial and Quantitative Analysis 42*, 857–892.

Negro, M. D. and C. Otrok (2008). Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. Staff reports, no. 326, Federal Reserve Bank of New York.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curve. *Journal of Business 60*, 473–489.

Ng, V. and Engle, R. F. and M. Rothschild (1992). A multi-dynamic factor model for stock returns. *Journal of Econometrics 52*, 245–266.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*, 1447–1479.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise. *Journal of Econometrics 168*, 244–258.

Otrok, C. and C. H. Whiteman (1998). Bayesian leading indicators: Measuring and predicting economic conditions in Iowa. *International Economic Review 39*, 997–1014.

Paisley, J. and L. Carin (ICML). Nonparametric factor analysis with beta process priors. *Annals of Applied Statistics*, 777–784.

Pati, D., A. Bhattacharya, N. S. Pillai, and D. B. Dunson (2012). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. Discussion Paper 2012-05, Department of Statistical Science, Duke University.

Peña, D. and G. E. P. Box (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association 82*, 836–843.

Peña, D. and P. Poncela (2004). Forecasting with nonstationary dynamic factor models. *Journal of Econometrics 119*, 291–321.

Pérez, J. M. and J. M. Berger (2002). Expected posterior prior distributions for model selection. *Biometrika 89*, 491–512.

Philipov, A. and M. E. Glickman (2006). Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics 24*, 313–328.

Pitt, M. and N. Shephard (1999). Time varying covariances: A factor stochastic volatility approach (with discussion). In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian statistics 6*, pp. 547–570. London: Oxford University Press.

Polasek, W. (1997). Factor analysis and outliers: A Bayesian approach. Discussion paper, University of Basel.

Press, S. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference (2nd edition)*. New York: Krieger.

Press, S. and K. Shigemasu (1989). Bayesian inference in factor analysis. In *ASA Proceedings of Social Statistics Section*, pp. 292–294.

Press, S. and K. Shigemasu (1994). Posterior distribution for the number of factors. In *ASA Proceedings of Bayesian Statistical Science*, pp. 75–77.

Rosenberg, B. and W. McKibben (1973). The prediction of systematic and specific risk in common stocks. *The Journal of Financial and Quantitative Analysis 8*, 17–333.

Rubin, D. B. and D. T. Thayer (1982). EM algorithms for ML factor analysis. *Psychometrika 47*, 69–76.

Rubin, D. B. and D. T. Thayer (1983). More on the EM for factor analysis. *Psychometrika 48*, 253–257.

Sahu, S. K. and K. V. Mardia (2005). A Bayesian kriged-Kalman model for short-term forecasting of air pollution levels. *Applied Statistics 54*, 223–44.

Sansó, B., A. M. Schmidt, and A. A. Nobre (2008). Bayesian spatio-temporal models based on discrete convolutions. *Canadian Journal of Statistics 36*, 239–58.

Sentana, E. (1998). The relation between conditionally heteroskedastic factor models and factor GARCH models. *Econometrics Journal 1*, 1–9.

Sentana, E., G. Calzolari, and G. Fiorentini (2008). Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks. *Journal of Econometrics 146*, 10–25.

So, M. K. P., K. Lam, and W. K. Li (1998). A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics 16*, 244–253.

Song, X. Y. and S. Y. Lee (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical & Statistical Psychology 54*, 237–263.

Spearman (1904). General intelligence objectively determined and measured. *American Journal of Psychology 15*, 201–292.

Stein, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging.* New York: Springer.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*, 147–162.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for var analysis. Working paper no. 11467, NBER.

Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 516–554. Elsevier.

Sun, D., R. Tsutakawa, and P. Speckman (1999). Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika 86*, 341–350.

Thomson, G. H. (1953). *The Factor Analysis of Human Ability.* University of London Press, London.

Thurstone, L. L. (1935). *Vectors of the Mind.* University of Chicago Press, Chicago.

Thurstone, L. L. (1947). *Multiple-Factor Analysis.* University of Chicago Press, Chicago.

Tsay, R. (2005). *Analysis of Financial Time Series.* Wiley.

Uhlig, H. and P. A. Ahmadi (2012). Measuring the dynamic effects of monetary policy shocks: A Bayesian FAVAR approach with sign restriction. 2012 Meeting Papers 1060, Society for Economic Dynamics.

Utsugi, A. and T. Kumagai (2001). Bayesian analysis of mixture of factor analysers. *Neural Computation 13*, 993–1002.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*, 177–188.

Vrontros, I. D., P. Dellaportas, and D. N. Politis (2003). A full-factor multivariate GARCH model. *Econometrics Journal 6*, 312–334.

Wang, F. and M. M. Wall (2003). Generalized common spatial factor model. *Biostatistics 4*, 569–582.

West, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 7*, pp. 723–732. London: Oxford University Press.

West, M. and P. J. Harrison (1997). *Bayesian Forecasting and Dynamic Model.* New York: Springer-Verlag.

Wikle, C. K. and N. Cressie (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika 86*, 815–829.

Yoshida, R. and M. West (2013). Bayesian learning in sparse graphical factor models via variational mean-field annealing. *Journal of Machine Learning Research 11*, 1771–1798.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.

Zhang, Z. and J. R. Nesselroade (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research 42*, 729.

Zhou, X., J. Nakajima, and M. West (2012). Bayesian forecasting and portfolio decisions using dynamic dependent factor models. Discussion Paper 2012-09, Department of Statistical Science, Duke University.