Statistica Sinica **14**(2004), 41-67

BAYESIAN MODEL ASSESSMENT IN FACTOR ANALYSIS

Hedibert Freitas Lopes and Mike West

Federal University of Rio de Janeiro and Duke University

Abstract: Factor analysis has been one of the most powerful and flexible tools for assessment of multivariate dependence and codependence. Loosely speaking, it could be argued that the origin of its success rests in its very exploratory nature, where various kinds of data-relationships amongst the variables at study can be iteratively verified and/or refuted. Bayesian inference in factor analytic models has received renewed attention in recent years, partly due to computational advances but also partly to applied focuses generating factor structures as exemplified by recent work in financial time series modeling. The focus of our current work is on exploring questions of uncertainty about the number of latent factors in a multivariate factor model, combined with methodological and computational issues of model specification and model fitting. We explore reversible jump MCMC methods that build on sets of parallel Gibbs sampling-based analyses to generate suitable empirical proposal distributions and that address the challenging problem of finding efficient proposals in high-dimensional models. Alternative MCMC methods based on bridge sampling are discussed, and these fully Bayesian MCMC approaches are compared with a collection of popular model selection methods in empirical studies. Various additional computational issues are discussed, including situations where prior information is scarce, and the methods are explored in studies of some simulated data sets and an econometric time series example.

Key words and phrases: Bayes factors, Bayesian inference, bridge sampling, expected posterior prior, latent factor models, model selection criteria, model uncertainty, reversible jump MCMC.

1. Introduction

Methodological innovations and real-world applications of factor analysis, and latent structure models more generally, have developed rapidly in recent years, partly due to increased access to appropriate computational tools. In particular, iterative MCMC simulation methods have very naturally opened up access to fully Bayesian treatments of factor analytic models, as developed and applied in, for example, Geweke and Zhou (1996), Polasek (1997), Arminger and Muthén (1998) and, with extensions to dynamic factor components in financial time series modelling (Aguilar and West (2000), Pitt and Shephard (1999)). The growing range of developments and creative applications in increasingly complex models, and with larger data sets in higher dimensions, justifies the view

that computational advances have been critically enabling; the near future will very likely see much broader use of factor analysis in routine applied statistical work. The above studies, and others, explore fully Bayesian inference in latent factor models in which the number of factors is a modelling choice; applied work typically studies sensitivity of predictions and variations/ambiguities of interpretations as the number of factors is varied as a control parameter. Formal inference on the number of factors itself has been relatively ignored in the Bayesian literature, though there are ranges of standard likelihood and frequentist methods available. Some key additional references, Bayesian and non-Bayesian, include (in order of appearance) Lawley and Maxwell (1963), Joreskog (1967), Martin and McDonald (1981), Bartholomew (1981), Press (1982, Chap.10), Lee (1981), Akaike (1987), Bartholomew (1987), Press and Shigemasu (1989) and Press and Shigemasu (1994). The book by Bartholomew (1987) is an excellent overview of the field up to about ten years ago.

The key issue of inference on, and selection of, the number of factors is the focus of this paper. Most recently, Polasek (1997) explored approaches to computing approximate posterior probabilities on the number of factors based on using MCMC methods for separate models differing only in the number of factors. Such an approach requires the computation of the observed values of the marginal data densities (prior predictive densities) under each of these separate models, for it is just these values that define the (marginal) likelihood function for inference on the number of factors, and the resulting Bayes' factors for pairwise model comparisons. This computation lies at the heart of the model selection and comparison problem. There are serious practical questions about choice and specification of prior distributions within the individual models, but that is not our primary focus here. A variety of methods are available for computing these marginal data density values – often referred to as the normalising constant problem. Some are specific to analyses based on MCMC methods within each individual model, and some are generic and based on analytic and asymptotic arguments. A wide ranging review of some standard methods appears in Kass and Raftery (1995), where the connections between various methods of approximating Bayes' factors using combinations of analytic and asymptotic arguments are well explored. These standard methods are closely related to non-Bayesian model selection criteria, including the well-known AIC, BIC/Schwartz criteria, and extensions of them using information-theoretic ideas, such as the ICOMP methods of Bozdogan and Ramirez (1987) and Bozdogan and Shigemasu (1998). Methods of approximating the marginal data densities that utilise outputs from MCMC analyses of separate models are of more interest here. Some of the methods we consider below are: the so-called candidate estimator (Chib (1995)), the harmonic mean estimator (Newton and Raftery (1994)), Gelfand and Dey's estimator (Gelfand and Dey (1994)), the Laplace-Metropolis estimator of Lewis

and Raftery (1997), and various novel approaches based on the recent innovative developments in bridge sampling (Meng and Wong (1996)). Additional useful references in this general area include, for example, Gilks, Richardson and Spiegelhalter (1996), DiCiccio, Kass, Raftery and Wasserman (1997) and Godsill (2001), which study comparisons and connections between some of the various methods just referenced.

Our paper has two main goals. First, we introduce, develop and explore 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 introduce a customised reversible jump Markov chain Monte Carlo (hereafter RJMCMC, see Green (1995)) algorithm for moving between models with different numbers of factors. RJMCMC approaches avoid the need for computing marginal data densities by treating the number of factors as a parameter, but require ingenuity in designing appropriate jumping rules to produce computationally efficient and theoretically effective methods. To compare with this, we introduce alternative methods based on bridge sampling ideas (Meng and Wong (1996)) that are specifically designed for computing the required marginal data densities in MCMC contexts, and represent the current frontiers of the field. Our second main goal is to explore these approaches and compare them with a range of standard model selection criteria and alternative methods of computing marginal data densities and Bayes' factors, as discussed above.

Section 2 defines the basic factor model framework, notation and structure, and discusses issues of model specification. Section 3 describes Bayesian analysis of the factor model when the number of factors is specified, based on standard Gibbs sampling. Section 4 describes the RJMCMC we introduce to address uncertainty about the number of factors. Section 5 briefly details some of the standard model selection criteria and alternative methods of marginal data density computation. Section 6 presents some comparative studies with simulated and real data sets. Section 7 concludes the paper with summary comments.

2. Factor Model Structure and Specification

2.1. Basic model form

Data on m related variables are considered to arise through random sampling from a zero-mean multivariate normal distribution denoted by $N(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega}$ denotes an $m \times m$ non-singular variance matrix. A random sample of size T is denoted by $\{\boldsymbol{y}_t, t = 1, \ldots, T\}$. For any specified positive integer $k \leq m$, the standard $k-$factor model relates each $\boldsymbol{y}_t$ to an underlying $k-$vector of random variables $\boldsymbol{f}_t$, the common factors, via

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

where (i) the factors $\boldsymbol{f}_t$ are independent with $\boldsymbol{f}_t \sim N(\boldsymbol{0}, \boldsymbol{I}_k)$, (ii) the $\boldsymbol{\epsilon}_t$ are independent normal $m-$vectors with $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \cdots, \sigma_m^2)$, (iii) $\boldsymbol{\epsilon}_t$ and $\boldsymbol{f}_s$ are independent for all t and s, and (iv) $\boldsymbol{\beta}$ is the $m \times k$ factor loadings matrix.

Under this model, the variance-covariance structure of the data distribution is constrained; we have $\boldsymbol{\Omega} = V(\boldsymbol{y}_t|\boldsymbol{\Omega}) = V(\boldsymbol{y}_t|\boldsymbol{\beta}, \boldsymbol{\Sigma})$ given by $\boldsymbol{\Omega} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}$. The model implies that, conditional on the common factors, the observable variables are uncorrelated: hence the common factors explain all the dependence structure among the m variables. For any elements y_{it} and y_{jt} of $\boldsymbol{y}_t$ and conditionally on $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, we have the characterising moments, (i) $\text{var}(y_{it}|\boldsymbol{f}) = \sigma_i^2$, (ii) $\text{cov}(y_{it}, y_{jt}|\boldsymbol{f}) = 0$, (iii) $\text{var}(y_{it}) = \sum_{l=1}^k \beta_{il}^2 + \sigma_i^2$, and (iv) $\text{cov}(y_{it}, y_{jt}) = \sum_{l=1}^k \beta_{il}\beta_{jl}$.

In practical problems, especially with larger values of m, the number of factors k will often be small relative to m, so that much of the variance-covariance structure is explained by the common factors. The *uniquenesses,* or *idiosyncratic variances,* σ_i^2 measure the residual variability in each of the data variables once that contributed by the factors is accounted for. The model (1) can be written as

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{\beta}' + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)'$, $\boldsymbol{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_T)'$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T)'$ are matrices of dimension $(T \times m), (T \times k)$ and $(T \times m)$, respectively. The elements $\boldsymbol{\epsilon}$ and $\boldsymbol{F}$ are mutually independent matrix normal random variables, as in Dawid (1981), Press (1982) and West and Harrison (1997, Chap. 16). The notation, as in Dawid (1981), is simply $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}_T, \boldsymbol{\Sigma})$. We then have densities

$$p(\boldsymbol{y}|\boldsymbol{F}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-T/2}\text{etr}(-0.5\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \tag{3}$$

and, marginalising over $\boldsymbol{F}$,

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Omega}|^{-T/2}\text{etr}(-0.5\boldsymbol{\Omega}^{-1}\boldsymbol{y}'\boldsymbol{y}), \tag{4}$$

where $\text{etr}(\mathbf{A}) = \exp(\text{trace}(\mathbf{A}))$ for any matrix $\mathbf{A}$. The likelihood function (3) will be subsequently used in Gibbs sampling for the parameters of a factor model with k fixed, whereas the likelihood form (4) will be extensively used in RJMCMC algorithms that also treat uncertainty about k.

2.2. Model structure and identification issues

As is well-known, the $k-$factor model must be further constrained to define a unique model free from identification problems. First we address the standard issue that the model is invariant under transformations of the form $\boldsymbol{\beta}^* = \boldsymbol{\beta}\boldsymbol{P}'$ and $\boldsymbol{f}_t^* = \boldsymbol{P}\boldsymbol{f}_t$, where $\boldsymbol{P}$ is any orthogonal $k \times k$ matrix. There are many ways

of identifying the model by imposing constraints on $\boldsymbol{\beta}$, including constraints to orthogonal $\boldsymbol{\beta}$ matrices, and constraints such that $\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}$ is diagonal (see Seber (1984)), for example). The alternative preferred here is to constrain so that $\boldsymbol{\beta}$ is a block lower triangular matrix, assumed to be of full rank, with diagonal elements strictly positive. This form is used, for example, in Geweke and Zhou (1996) and Aguilar and West (2000), and provides both identification and, often, useful interpretation of the factor model. In this form, the loadings matrix has $r = mk - k(k-1)/2$ free parameters. With m non-zero σ_i parameters, the resulting factor form of $\boldsymbol{\Omega}$ has $m(k+1) - k(k-1)/2$ parameters, compared with the total $m(m+1)/2$ in an unconstrained (or $k = m$) model; leading to the constraint that $m(m+1)/2 - m(k+1) + k(k-1)/2 \geq 0$, which provides an upper bound on k. For example, $m = 6$ implies $k \leq 3$, $m = 12$ implies $k \leq 7$, $m = 20$ implies $k \leq 14$, $m = 50$ implies $k \leq 40$, and so on. Even for small m, the bound will often not matter as relevant k values will not be so large. In realistic problems, with m 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 setting one or more of the residual variances σ_i to zero.

A question arises about the full-rank assumption for $\boldsymbol{\beta}$. This was addressed in Geweke and Singleton (1980) who show that, if $\boldsymbol{\beta}$ is rank deficient, then the model is unidentified. Specifically, if $\boldsymbol{\beta}$ has rank $r < k$ there exists a matrix $\boldsymbol{Q}$ such that $\boldsymbol{\beta Q} = \boldsymbol{0}$, $\boldsymbol{Q'Q} = \boldsymbol{I}$ and, for any orthogonal matrix $\boldsymbol{M}$,

$$\boldsymbol{\Omega} = \boldsymbol{\beta\beta}' + \boldsymbol{\Sigma} = (\boldsymbol{\beta} + \boldsymbol{MQ}')'(\boldsymbol{\beta} + \boldsymbol{MQ}') + (\boldsymbol{\Sigma} - \boldsymbol{MM}'). \tag{5}$$

This translation invariance of $\boldsymbol{\Omega}$ under the factor model implies lack of identification and, in application, induces symmetries and potential multimodalities in resulting likelihood functions. This issue relates intimately to the question of uncertainty of the number of factors, discussed further below.

A final question concerns the ordering of the y_{it} variables and the connection between a chosen ordering and the specific form of the factor loading matrix above. The order of variables is a modelling decision that has no effect on the resulting theoretical model nor on predictive inferences under the model. Given the k-factor model (1) specified and appropriate for the $\boldsymbol{y}$ with variables in a specific order, alternative orderings are trivially produced via $\boldsymbol{Ay}_t$ for some rotation matrix $\boldsymbol{A}$. Model (1) then transforms to a similar factor model for the reordered data $\boldsymbol{Ay}_t$ with the same latent factors but transformed loadings matrix $\boldsymbol{A\beta}$. This new loadings matrix does not have the lower triangular structure. However, we can always find an orthonormal matrix $\boldsymbol{P}$ such that $\boldsymbol{A\beta P}'$ is lower triangular, and so simply recover the factor model in precisely the form (1) with the same probability structure for the underlying latent factors $\boldsymbol{Pf}_t$. This result confirms that the order of the variables in $\boldsymbol{y}_t$ is theoretically irrelevant assuming that k

is properly chosen. However, when it comes to model estimation, the order of variables has a determining effect on the choice of k, and the interaction between variable order and model fitting can be quite subtle, as we illustrate in examples below.

2.3. Elements of prior specification

To complete the model specification we require classes of priors for the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. Our reported analyses are based on very diffuse but proper priors with the following ingredients. For the factor loadings, we take independent priors such that $\beta_{ij} \sim N(0, C_0)$ when $i \neq j$, and $\beta_{ii} \sim N(0, C_0)\mathbf{1}(\beta_{ii} > 0)$ for the upper-diagonal elements of positive loadings $i = 1, \cdots, k$. The latter simply truncates the basic normal prior to restrict the diagonal elements to positive values. Analysis now requires only that we specify the variance parameter C_0, which we take to be rather large in the studies below.

For each of the idiosyncratic variances σ_i^2 we assume a common inverse gamma prior, and take the variances to be independent. Specifically, the σ_i^2 are independently modelled as $\sigma_i^2 \sim IG(\nu/2, \nu s^2/2)$ with specified hyperparameters ν and s^2. Here s^2 is the prior mode of each σ_i^2 and ν the prior degrees of freedom hyperparameter. Our examples below assume low values of ν to produce diffuse though proper priors. Note that we eschew the use of standard improper reference priors $p(\sigma_i^2) \propto 1/\sigma_i^2$. Such priors lead to the Bayesian analogue of the so-called *Heywood problem* (Martin and McDonald (1981), Ihara and Kano (1995)). In terms of these variance parameters, likelihood functions in factor models are bounded below away from zero as σ_i^2 tends to zero, so inducing singularities in the posterior at zero. Proper priors that decay to zero at the origin obviate this problem and induce proper posteriors.

3. MCMC Methods in a $k-$factor Model

With a specified $k-$factor model, Bayesian analyses using MCMC methods are straightforward. We simply summarise the main ingredients here, referring to Geweke and Zhou (1996), Polasek (1997) and Aguilar and West (2000) for further details. MCMC analysis involves iteratively simulating from sets of conditional posterior distributions which, in this model, are standard forms. A basic method simulates from the conditional posteriors for each of $\boldsymbol{F}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in turn, utilising the following sets of full conditional posteriors arising from our model as specified. These are as follows.

First, the factor model in (2) can be seen as a standard multivariate regression model with "parameters" $\boldsymbol{F}$ when $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and k are fixed (e.g., Press (1982), Box and Tiao (1973), Broemeling (1985) and Zellner (1971)). It easily

follows that the full conditional posterior for $\boldsymbol{F}$ factors into independent normal distributions for the $\boldsymbol{f}_t$, namely

$$\boldsymbol{f}_t \sim N((\boldsymbol{I}_k + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t, (\boldsymbol{I}_k + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1})$$

independently for $t = 1, \ldots .T$.

Second, the full conditional posterior for $\boldsymbol{\beta}$ also factors into independent margins for the non-zero elements of the rows of $\boldsymbol{\beta}$, as follows. For rows $i = 1, \ldots, k$, write $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ii})'$ for just these non-zero elements. For the remaining rows $i = k+1, \ldots, m$, write $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ik})$. Similarly, for $i = 1, \ldots, k$ denote by $\boldsymbol{F}_i$ the $T \times i$ matrix containing the first i columns of $\boldsymbol{F}$, and for all i let $\boldsymbol{y}_i$ be the column i of $\boldsymbol{y}$.

Finally, it is trivially deduced that full conditional posterior for the elements of $\boldsymbol{\Sigma}$ reduces to a set of m independent inverse gammas, with $\sigma_i^2 \sim IG((\nu + T)/2, (\nu s^2 + d_i)/2)$ where $d_i = (\boldsymbol{y}_i - \boldsymbol{F}\boldsymbol{\beta}_i')'(\boldsymbol{y}_i - \boldsymbol{F}\boldsymbol{\beta}_i')$.

Then we have full conditionals as follows:

- for $i = 1, \ldots, k$, $\boldsymbol{\beta}_i \sim N(\boldsymbol{m}_i, \boldsymbol{C}_i)\mathbf{1}(\beta_{ii} > 0)$ where $\boldsymbol{m}_i = \boldsymbol{C}_i(C_0^{-1}\mu_0\mathbf{1}_i + \sigma_i^{-2}\boldsymbol{F}_i'\boldsymbol{y}_i)$ and $\boldsymbol{C}_i^{-1} = C_0^{-1}\boldsymbol{I}_i + \sigma_i^{-2}\boldsymbol{F}_i'\boldsymbol{F}_i$;
- for $i = k+1, \ldots, m$, $\boldsymbol{\beta}_i \sim N(\boldsymbol{m}_i, \boldsymbol{C}_i)$ where $\boldsymbol{m}_i = \boldsymbol{C}_i(C_0^{-1}\mu_0\mathbf{1}_k + \sigma_i^{-2}\boldsymbol{F}'\boldsymbol{y}_i)$ and $\boldsymbol{C}_i^{-1} = C_0^{-1}\boldsymbol{I}_k + \sigma_i^{-2}\boldsymbol{F}'\boldsymbol{F}$.

These distributions are easily simulated.

4. Fully Bayesian Inference on the Number of Factors

4.1. Preliminary parallel MCMC analyses

The above section provides the basis for posterior simulations in a model with k specified. Reversible jump MCMC (RJMCMC) methods are useful for exploring posterior distributions for model parameters in the context of uncertainty about k, and with k included as a parameter. As we move between models with different numbers of factors, the dimension and meaning of the model parameters change, and RJMCMC methods are designed for just such problems.

For this and the following sections, we make explicit the dependence of the factor loading matrix on k by refining the notation, replacing $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_k$ and $\boldsymbol{F}$ by $\boldsymbol{F}_k$. Further, we write $\boldsymbol{\theta}_k$ for the parameters $(\boldsymbol{\beta}_k, \boldsymbol{\Sigma})$ of a $k-$factor model. The number k now appears explicitly in the conditioning of all model density functions. Our RJMCMC methods involve Metropolis-Hastings type algorithms that move a simulation analysis between models defined by $(k, \boldsymbol{\theta}_k)$ to $(k', \boldsymbol{\theta}_{k'})$ with different defining dimensions k and k'. The resulting Markov chain simulations jump between such distinct models, and the algorithms are designed to be reversible so as to maintain detailed balance of the chain. Futher details of the general methodology and ideas can be found in Green (1995).

Our method builds on a preliminary set of parallel MCMC analyses that are run over a set of prespecified values $k \in K$ of the number of factors. These chains produce a set of K within-model posterior samples for $(\boldsymbol{\theta}_k, \boldsymbol{F}_k)$ that approximate the posterior distributions $p(\boldsymbol{\theta}_k, \boldsymbol{F}_k|k, \boldsymbol{y})$. From these samples we compute posterior means and other summaries, and use these to guide choice of analytically specified distributions to be used to generate candidate parameter values in the RJMCMC algorithm. This component of the analysis operates only with the samples for the parameters $\boldsymbol{\theta}_k$, the simulated values of the actual factors $\boldsymbol{F}_k$ being relevant but incidental to the moving between models with different values of k. Write $\boldsymbol{b}_k$ and $\boldsymbol{B}_k$ for the approximate posterior mean and variance matrix of $\boldsymbol{\beta}_k$ from the MCMC analysis and, for each $i = 1, \ldots, m$, write v_{ki}^2 for the approximate posterior mode of σ_i^2 from the analysis. In our current implementation we introduce the following analytic forms as components of a proposal distribution. For each model order $k \in K$, $q_k(\boldsymbol{\beta}_k) = N(\boldsymbol{b}_k, b\boldsymbol{B}_k)$ and for $i = 1, \ldots, m$, $q_k(\sigma_i^2) = IG(a, av_{ki}^2)$, where a and b are positive scale parameters to be specified. These density functions are combined to produce the distributions

$$q_k(\boldsymbol{\theta}_k) \equiv q_k(\boldsymbol{\beta}_k, \boldsymbol{\Sigma}) = q_k(\boldsymbol{\beta}_k) \prod_{i=1}^{m} q_k(\sigma_i^2), \qquad k \in K \tag{6}$$

for use as now described.

4.2. A reversible jump algorithm

Following the set of preliminary MCMC analyses for models in parallel, we explore the space of models as k varies using the following version of RJMCMC. In addition to the $k-$factor models and within model priors specified above, we need to specify the marginal prior probabilities $p(k)$ over $k \in K$. Then the RJMCMC analysis proceeds as follows.

0. Choose a starting value of k. Set the current values of $\boldsymbol{\theta}_k$ to a draw from the posterior $p(\boldsymbol{\theta}_k|k, \boldsymbol{y})$ by using one (or more) steps of the MCMC algorithm as described above and based on past sampled values from this $k-$factor model. Note that this step produces both new sampled values of $\boldsymbol{\theta}_k$ and the factors $\boldsymbol{F}_k$, though only the former are used in exploring moves to models with other k values.

1. *Between model* move step:

1.a Draw a candidate value of the number of factors k' from a proposal distribution defined by prespecified transition probabilities $Pr(k'|k) = J(k \to k')$.

1.b Given k', draw the parameters $\boldsymbol{\theta}_{k'}$ from the distribution $q_{k'}(\boldsymbol{\theta}_{k'})$ of equation (6).

1.c Compute the accept/reject ratio

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{y}|k', \boldsymbol{\theta}_{k'})p(\boldsymbol{\theta}_{k'}|k')p(k')}{p(\boldsymbol{y}|k, \boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k)p(k)} \frac{q_k(\boldsymbol{\theta}_k)J(k' \to k)}{q_{k'}(\boldsymbol{\theta}_{k'})J(k \to k')}\right\}. \quad (7)$$

Here, for each value $j \in (k, k')$, $p(\boldsymbol{y}|j, \boldsymbol{\theta}_j) = p(\boldsymbol{y}|j, \boldsymbol{\beta}_j, \boldsymbol{\Sigma})$ is the likelihood function in equation (4), $p(\boldsymbol{\theta}_j|j)$ is the prior density function for the parameters within the $j-$factor model, and $p(j)$ is the prior probability on k factors. With probability α, accept the jump to the $k'-$factor model and the new parameter values $\boldsymbol{\theta}_{k'}$ just sampled as candidates.

2. *Within model* move step:

If the jump to model k' is accepted, run one step of the MCMC analysis in this $k'-$factor model, producing new sample values of the full set of quantities $(\boldsymbol{\theta}_{k'}, \boldsymbol{F}_{k'})$. Otherwise, remain in model k and use the MCMC to produce new values of $(\boldsymbol{\theta}_k, \boldsymbol{F}_k)$.

3. Repeat [1] and [2] until practical convergence is judged to have been achieved.

The chosen proposal distributions $q_k(\boldsymbol{\theta}_k)$ are not generally expected to provide globally accurate approximations to the conditional posteriors $p(\boldsymbol{\theta}_k|k, \boldsymbol{y})$. However, if that happened to be the case then the resulting accept/reject probabilities above reduce directly to Metropolis-type probabilities on the parameter k alone. Our algorithm is a particular case of what Dellaportas, Forster and Ntzoufras (2002) called the *Metropolised Carlin and Chib* method, where the proposal distributions generating both new model dimension and new parameters depend on the current state of the chain only through k. This is true here as we use proposals based on the initial, auxilliary MCMC analyses. A more descriptive name is independence RJMCMC, analogous to the standard terminology for independence Metropolis-Hastings methods.

5. Other Methods of Addressing Model Uncertainty

5.1. Introductory comments

The RJMCMC technology is becoming standard in Bayesian work with competing models with differing numbers of parameters. By comparison with traditional approaches based on nesting models in a "super-model", RJMCMC is often more efficient computationally and, in terms of practical computing time

to convergence of the Markov chains, it has an established theory that guarantees convergence of the chains in very general frameworks. Thus, in addition to having a very natural and direct specification in our factor model context, convergence of the chain to sampling from the full posterior across models, as well as for parameters and factors within models, is ensured. There are, however, ranges of existing methods for approximate inference on the number of factors, and we aim to compare these methods in examples below. This section provides a brief catalogue description of methods and model selection criteria, as well as introducing a novel approach based on recent work in bridge sampling.

5.2. Computing normalizing constants

In our Bayesian framework the within-model analysis provides, in theory, the marginal data density functions

$$p(\boldsymbol{y}|k) = \int p(\boldsymbol{y}|k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|k) d\boldsymbol{\theta}_k \tag{8}$$

for each value of $k \in K$. If these could be computed, then inference on k follows from Bayes' theorem via $p(k|\boldsymbol{y}) \propto p(k)p(\boldsymbol{y}|k)$. The problem is computational: the marginal data densities are generally not easily computed and so must be approximated numerically. The following standard methods are of interest.

5.2.1. Candidate's estimator

The so-called candidate's estimator, first referred in Besag (1989) and fully analysed by Chib (1995), is of interest when the $k-$factor models are each analysed using MCMC. The approach observes that, for *any* value of $\boldsymbol{\theta}_k$, Bayes' theorem implies that

$$p(\boldsymbol{y}|k) = \frac{p(\boldsymbol{y}|k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|k)}{p(\boldsymbol{\theta}_k|k, \boldsymbol{y})}. \tag{9}$$

The idea is then to estimate the components of this equation that are not available analytically, then plug-in a chosen value of $\boldsymbol{\theta}_k$ to provide an estimate of $p(\boldsymbol{y}|k)$. Our analysis uses the posterior means from the MCMC analyses as these plug-in values.

Now, the numerator in (9) factors as $p(\boldsymbol{y}|k, \boldsymbol{\beta}_k, \boldsymbol{\Sigma}) p(\boldsymbol{\beta}_k|k) p(\boldsymbol{\Sigma}|k)$ each component of which can be directly and easily evaluated. The posterior density function in the denominator requires approximation, and with some creativity. Theoretically, this is given by

$$p(\boldsymbol{\beta}_k, \boldsymbol{\Sigma}|k, \boldsymbol{y}) = p(\boldsymbol{\beta}_k|k, \boldsymbol{y}) p(\boldsymbol{\Sigma}|k, \boldsymbol{\beta}_k, \boldsymbol{y}), \tag{10}$$

where the two terms in the right hand side are approximated, respectively by:

$$p(\boldsymbol{\beta}_k|k, \boldsymbol{y}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\boldsymbol{\beta}_k|k, \boldsymbol{\Sigma}^{(m)}, \boldsymbol{F}_k^{(m)}, \boldsymbol{y}), \tag{11}$$

$$p(\boldsymbol{\Sigma}|k,\boldsymbol{\beta}_k,\boldsymbol{y}) \approx \frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{\Sigma}|k,\boldsymbol{\beta}_k,\boldsymbol{F}_{k1}^{(m)},\boldsymbol{y}), \tag{12}$$

where the sum, in the first approximation, is over draws $(\boldsymbol{F}_k^{(m)},\boldsymbol{\Sigma}^{(m)})$ from the MCMC analysis and can be easily performed, since it is a sum of multivariate normal distributions (truncated).

The second approximation, however, is a sum (of products of inverse gammas) over draws $\boldsymbol{F}_{k1}^{(m)}$ from an MCMC analysis conditional on $\boldsymbol{\beta}_k$. To do this we need to run additional MCMC chains in each $k-$factor model with the $\boldsymbol{\beta}_k$ fixed at its chosen value. This naturally introduces a significant additional computational burden, especially in larger models.

In studies below we denote by $\hat{p}_C$ the resulting approximation to $p(\boldsymbol{y}|k)$.

5.2.2. Harmonic mean estimator

In a similar spirit to the candidate's estimator, the harmonic mean estimator makes use of the identity

$$p(\boldsymbol{y}|k)^{-1} = \int p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)^{-1} p(\boldsymbol{\theta}_k|k,\boldsymbol{y}) d\boldsymbol{\theta}_k. \tag{13}$$

As discussed in Newton and Raftery (1994), the resulting estimator is based on the importance sampling approximation to the integral using the exact posterior as importance sampling distribution. This results in the approximation $p(\boldsymbol{y}|k) \approx \hat{p}_H$ where

$$\hat{p}_H^{-1} = M^{-1}\sum_{m=1}^{M} p(\boldsymbol{y}|k,\boldsymbol{\theta}_k^{(m)})^{-1}, \tag{14}$$

where the $\boldsymbol{\theta}_k^{(m)}$ are posterior samples from the MCMC analysis and the density evaluations are made using equation (4). Note that $\hat{p}_H$ is an harmonic mean of the likelihood values, hence the name. Newton and Raftery (1994) discuss the accuracy of $\hat{p}_H$ among other issues. Though it has been quite widely used, it can be unstable in some applications since small likelihood values can overly influence the resulting harmonic mean value.

5.2.3. Newton and Raftery's estimator

Partly motivated by the stability issues associated with $\hat{p}_H$, Newton and Raftery (1994) suggested estimators defined as follows. Let $g(\boldsymbol{\theta}_k) = \delta p(\boldsymbol{\theta}_k|k) + (1-\delta)p(\boldsymbol{\theta}_k|k,\boldsymbol{y})$ be a mixture of the prior and posterior for $\boldsymbol{\theta}_k$ for some small mixing probability δ. Sampling from $g(\cdot)$ is easy – simply randomly replace values in the available posterior sample by independent draws from the prior. Do this

iteratively, repeatedly computing the sequence of γ values defined by $\gamma^{new} = A(\gamma^{old})/B(\gamma^{old})$, where

$$A(\gamma) = \sum_{m=1}^{M} f_m\{\delta\gamma + (1-\delta)f_m\}^{-1} \text{ and } B(\gamma) = \sum_{m=1}^{M} \{\delta\gamma + (1-\delta)f_m\}^{-1}$$

and the quantities f_m are the likelihood values $f_m = p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)})$ evaluated at the current sample of M points $\boldsymbol{\theta}_k^{(m)}$ from the mixture $g(\cdot)$. Iterations of this procedure lead to a stable limiting value $\gamma = \hat{p}_{NR}$ that provides the proposed estimator of $p(\boldsymbol{y}|k)$. Notice that the case $\delta = 0$ implies that $\hat{p}_{NR} = \hat{p}_H$. Implementation of this method depends on the choice of δ.

The main problem with this estimator is that is depends on draws from both the posterior and the prior. Newton and Raftery (1994) have suggested combining the M draws from the posterior with $\delta M/(1-\delta)$ draws from the the prior, all of them with the same likelihood equal to their expected value, $p(\boldsymbol{y}|K)$, the predictive density. Once again, the solution is found iteratively as, $\gamma^{new} = A(\gamma^{old})/B(\gamma^{old})$, where

$$A(\gamma) = \epsilon M + \sum_{m=1}^{M} f_m\{\delta\gamma + (1-\delta)f_m\}^{-1} \text{ and } B(\gamma) = \epsilon M\gamma^{-1} + \sum_{m=1}^{M} \{\delta\gamma + (1-\delta)f_m\}^{-1}$$

with $\epsilon = \delta/(1-\delta)$.

5.2.4. Gelfand and Dey's estimator

Related estimators introduced by Gelfand and Dey (1994) are inspired by the identities

$$p(\boldsymbol{y}|k)^{-1} = \int g(\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k, \boldsymbol{y})\{p(\boldsymbol{\theta}_k|k)p(\boldsymbol{y}|k, \boldsymbol{\theta}_k)\}^{-1}d\boldsymbol{\theta}_k \tag{15}$$

that hold for all arbitrary densities $g(\cdot)$. These authors study approximations $\hat{p}_{GD}$ given by

$$\hat{p}_{GD}^{-1} = M^{-1} \sum_{m=1}^{M} g(\boldsymbol{\theta}_k^{(m)})\{p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)})p(\boldsymbol{\theta}_k^{(m)}|k)\}^{-1}, \tag{16}$$

where, again, the $\boldsymbol{\theta}_k^{(m)}$ are posterior samples. The $\hat{p}_{GD}$ is unstable if g has thicker tails than $p(\boldsymbol{y}|k, \boldsymbol{\theta})p(\boldsymbol{\theta}_k|k)$.

5.2.5 Laplace-Metropolis estimator

The Laplace-Metropolis estimator combines analytic posterior approximations with MCMC output to modify traditional Laplace asymptotics (Tierney

and Kadane (1986)). Discussed in Lewis and Raftery (1997), the resulting estimator has the form

$$\hat{p}_{LM} = (2\pi)^{d/2}|\boldsymbol{\Psi}|^{1/2}p(\boldsymbol{y}|k,\tilde{\boldsymbol{\theta}}_k)p(\tilde{\boldsymbol{\theta}}_k|k), \tag{17}$$

where $\tilde{\boldsymbol{\theta}}_k$ maximises $p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k)$ among the M posterior draws, $\boldsymbol{\Psi}$ is the MCMC approximation to the posterior variance of $\boldsymbol{\theta}_k$, and d is the dimension of $\boldsymbol{\theta}_k$. Variations on this method replace $\tilde{\boldsymbol{\theta}}_k$ by the MCMC approximation to the posterior mean.

5.2.6. Bridge sampling

Innovative methods based on bridge sampling have recently been studied by Meng and Wong (1996). In our context this applies as follows. If $g(\boldsymbol{\theta}_k)$ is any chosen proposal density function with the same support as the posterior $p(\boldsymbol{\theta}_k|k,\boldsymbol{y})$, note the identity $p(\boldsymbol{y}|k) = N/D$, where

$$N = \int \alpha(\boldsymbol{\theta}_k)g(\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k,$$
$$D = \int \alpha(\boldsymbol{\theta}_k)g(\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k,\boldsymbol{y})d\boldsymbol{\theta}_k.$$

Based on the MCMC sample values $\boldsymbol{\theta}_k^{(m)}$ from the posterior distribution and on L values $\boldsymbol{\theta}_k^{*(l)}$ from $g(\cdot)$ (an importance function) we now have an easy approximation $p(\boldsymbol{y}|k) \approx \hat{N}/\hat{D}$, where

$$\hat{N} = L^{-1}\sum_{l=1}^{L}\alpha(\boldsymbol{\theta}_k^{*(l)})p(\boldsymbol{\theta}_k^{*(l)}|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k^{*(l)}) \quad \text{and} \quad \hat{D} = M^{-1}\sum_{m=1}^{M}\alpha(\boldsymbol{\theta}_k^{(m)})g(\boldsymbol{\theta}_k^{(m)}).$$

Generally, we aim to choose $g(\cdot)$ to be as accurate an approximation to the posterior as possible, while remaining easy to compute and simulate.

Different choices of the "arbitrary" function $\alpha(\cdot)$ define different bridge sampling estimators. Some discussed by Meng and Wong (1996) are as follows.

- If $\alpha(\boldsymbol{\theta}_k) = (p(\boldsymbol{\theta}_k|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)g(\boldsymbol{\theta}_k))^{-1}$ the corresponding estimator resembles the harmonic mean estimator. We label this $\hat{p}_{GH}$ and note that it is obtained by letting

$$\hat{N} = L^{-1}\sum_{l=1}^{L} g(\boldsymbol{\theta}_k^{*(l)})^{-1} \quad \text{and} \quad \hat{D} = M^{-1}\sum_{l=1}^{M}\{p(\boldsymbol{\theta}_k^{(m)}|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k^{(m)})\}^{-1}.$$

- If $\alpha = (p(\boldsymbol{\theta}_k|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)g(\boldsymbol{\theta}_k))^{-1/2}$ we have what is called the geometric estimator $\hat{p}_G$. This is given by

$$\hat{N} = L^{-1}\sum_{l=1}^{L}\{p(\boldsymbol{\theta}_k^{*(l)}|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k^{*(l)})/g(\boldsymbol{\theta}_k^{*(l)})\}^{1/2},$$

$$\hat{D} = M^{-1} \sum_{l=1}^{M} \{ g(\boldsymbol{\theta}_k^{(m)}) / (p(\boldsymbol{\theta}_k^{(m)}|k) p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)})) \}^{1/2}.$$

- The *optimal estimator* of Meng and Wong (1996), denoted by $\hat{p}_{opt}$, is obtained by an iterative procedure. Specify the initial value $r^{\cdot} = \hat{p}_G$ and, defining $s_1 = 1 - s_2 = M/(M+L)$, iterate the equation $r^{new} = A(r^{old})/B(r^{old})$, where

$$A(r) = \sum_{l=1}^{L} W_{2i}/(s_1 W_{2l} + s_2 r) \quad \text{and} \quad B(r) = \sum_{m=1}^{M} 1/(s_1 W_{1m} + s_2 r),$$

$W_{2l} = p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{*(l)}) p(\boldsymbol{\theta}_k^{*(l)}|k) / g(\boldsymbol{\theta}_k^{*(l)})$ and $W_{1m} = p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)}) p(\boldsymbol{\theta}_k^{(m)}|k) / g(\boldsymbol{\theta}_k^{(m)})$, for $l = 1, \ldots, L$, and $m = 1, \ldots, M$, respectively.

5.3. Likelihood and information criteria

Traditional model selection criteria based on the likelihood include variants of AIC, Akaike (1987), the Schwartz or Bayesian criteria, or BIC, and related information criteria such as the ICOMP methods of Bozdogan and Ramirez (1987) and Bozdogan and Shigemasu (1998). Explicit equations for some of these criteria, that we use below in comparative studies, are provided here. For each $k-$factor model, write $l_k = -2\log(p(\boldsymbol{y}|k, \hat{\boldsymbol{\theta}}_k))$ where $\hat{\boldsymbol{\theta}}_k$ is MLE of $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\Sigma})$ and the likelihood function is the standard form in equation (4). Write $\hat{\boldsymbol{\Omega}}_k = \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k' + \hat{\boldsymbol{\Sigma}}$ for the corresponding MLE of $\boldsymbol{\Omega}_k$. It is easily deduced that

$$l_k = T \left\{ m \log(2\pi) + \log |\hat{\boldsymbol{\Omega}}_k| + \text{trace}(\hat{\boldsymbol{\Omega}}_k^{-1} \boldsymbol{S}) \right\},$$

where $\boldsymbol{S} = \boldsymbol{y}'\boldsymbol{y}/T$. The various model selection criteria are defined as follows: (i) AIC $= l_k + 2p_k$, (ii) BIC $= l_k + \log(T)p_k$, (iii) BIC $^* = l_k + \log(\tilde{T})p_k$, and (iv) ICOMP $= l_k + C_1(\hat{\boldsymbol{\Sigma}}_k)$, where $p_k = m(k+1) - k(k-1)/2, \tilde{T} = T - (2m + 11)/6 - 2k/3$ and $C_1(\boldsymbol{\Sigma}_k) = 2(k+1)\left((m/2)\log(\text{trace}\boldsymbol{\Sigma}_k/m) - 0.5\log|\boldsymbol{\Sigma}_k|\right)$. See Bozdogan and Shigemasu (1998) for further details.

6. Empirical Exploration and Comparison

6.1. A first simulation study

An initial simulation study considers a one-factor model for a seven-dimensional problem generating one hundred observations; thus $m = 7, k = 1$ and $T = 100$. In each of a series of simulations, T observations were drawn from a one-factor models defined by parameters

$$\beta' = (0.995, 0.975, 0.949, 0.922, 0.894, 0.866, 0.837),$$
$$diag(\Sigma) = (0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30).$$

In this first simulation study, and in the following ones, we have decided to model correlation matrices as opposed to covariance matrices. As extensively discussed in Section 2.2, the results are invariant to standardization of the variables. Besides, prior elicitation should be, at least in principle, more direct when dealing with correlations than with covariances.

Each such simulated data set was analysed using the MCMC and reversible jump methodologies, and also subject to study using the range of model selected criteria and methods described above. This study explored $k-$factor models for each data set, with up to three possible factors in each case.

MCMC analyses utilised the prior distributions based on the following hyperparameter values: for $\boldsymbol{\beta}$, hyperparameters $\mu_0 = 0$ and $C_0 = 1$ define very vague priors; for the variances σ_k^2, $\nu_{0i} = 2.2$ and $\nu_{0i}s_{0i}^2 = 0.1$ define similarly vague priors with prior means of 0.5 for each σ_k^2. The MCMC and reversible jump samplers were based on $M_0 = 10,000$ iterations as burn-in, followed by a further 10,000 iterates that were sampled every ten steps to produce a final MCMC sample of size 1,000. In generating proposals in the RJMCMC methods, we adopted $a = 18$, $b = 2$ and

$$J = \begin{pmatrix} 0.0 & 1.0 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.0 & 1.0 & 0.0 \end{pmatrix}.$$

Both a and b are also used to define the proposals, $g(\boldsymbol{\theta})$, for the other methods of addressing model uncertainty (see Section 5). Among the candidate methods for model selected, the "Newton and Raftery" technique requires the specification of a control parameter, δ; this was set at $\delta = 0.05$, and the number of iterations at 1,000. Table 1 displays results from this simulation analysis. We repeated the model fitting exercises for 1,000 different data sets generated independently from the one-factor model.

The table provides simple counts of the number of times that each k-factor model achieved the highest posterior probability. For example, the harmonic estimator method selected the one-factor model 428 times out of 1,000, and the three-factor model 314 times out of 1,000. Evidently, most of the approximation methods are very reliable in favouring the one-factor model, as is the RJMCMC (the "gold standard") approach. Bridge sampling methods agree with our RJMCMC approach. Relatively poor results are achieved by the harmonic mean method, Newton-Raftery estimator, AIC, ICOMP, and to some extent by the candidate's estimator, which all tend to prefer higher numbers of factors a significant proportion of the time. In terms of model selection *per se*, as opposed to exploring model uncertainty more formally, the BIC methods are relatively accurate and, of course, rather easier to compute.

Table 1. Comparison of model uncertainty assessment methods on simulated data set 1.

	k		
Method	$k=1$	$k=2$	$k=3$
RJMCMC	1000	0	0
$\hat{p}_C$	954	46	0
$\hat{p}_H$	428	258	314
$\hat{p}_{NR}$	467	234	299
$\hat{p}_{GD}$	1000	0	0
$\hat{p}_{LM}$	1000	0	0
$\hat{p}_G$	1000	0	0
$\hat{p}_{opt}$	1000	0	0
Criterion	$k=1$	$k=2$	$k=3$
AIC	854	135	11
BIC	1000	0	0
BIC*	1000	0	0
ICOMP	607	296	97

In analysis of real data, we run into multi-modalities in posterior distributions that require some thought and explanation. In anticipation of this, we here explore some summaries of a three-factor model fitted to one of the simulated data sets arising from this true one-factor model.

Figure 1 displays marginal posterior densities of some of the idiosyncratic variances from such an analysis. Note the multi-modality; marginal posteriors for elements of the $\boldsymbol{\beta}$ matrix exhibit corresponding multiple modes. This arises due to the mis-match between the model assumption of $k=3$ and the data structure based on $k=1$, and is driven by the identification issues arising in such cases, as Discussed in Section 2.2 around equation (5). Encountering such multi-modality in posteriors from a specified model can therefore be taken as a suggestion that the chosen value of k is too large.

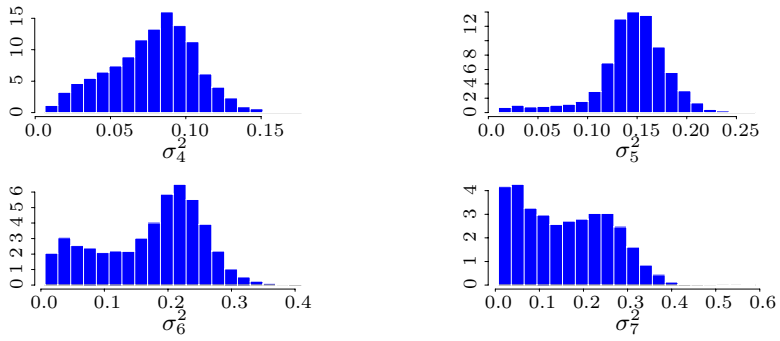

Figure 1. Marginal posteriors of the σ_i^2 $(i=4,\ldots,7)$ from analysis of the simulated data set from a one-factor structure but analysed using a model with $k=3$ factors (based on the first simulated study at Section 6.1).

6.2. A second simulation study

A second study follows the pattern of the above example, but now using data sets simulated from a model with $m = 9$ variables, $k = 3$ factors, and with a sample size of just $T = 50$. The true model in this case has parameters

$$\beta' = \begin{pmatrix} 0.99 & 0.00 & 0.00 & 0.99 & 0.99 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.95 & 0.00 & 0.00 & 0.00 & 0.95 & 0.95 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.90 & 0.00 & 0.00 & 0.00 & 0.00 & 0.90 & 0.90 \end{pmatrix},$$

$$\mathrm{diag}(\Sigma) = (0.02, 0.19, 0.36, 0.02, 0.02, 0.19, 0.19, 0.36, 0.36).$$

The analyses used the same hyper-parameters and MC sample size choices as in the first simulation study. As there, we summarise one aspect of performance of model selection methods by simply counting the number of times, out of a total of 1,000 analyses of simulated data sets, that each possible k-factor model received the highest posterior probability using each of the methods of computation. These summaries appear in Table 2. Again, it is clear that several of the approximation methods reliably identify the true model structure, which gives some indication of their likely utility in real data analyses. Among the approximate Bayesian methods, those based on the candidate's estimator, the harmonic mean estimator and the Newton-Raftery technique are the only failures, their performances being quite poor in comparison to the other Bayesian approaches and to the information criteria. Once again, AIC and ICOMP choose the wrong model in at least 12 percent of the simulations. AIC is often known to overestimate the size of the model, as empirically observed here.

Table 2. Comparison of model uncertainty assessment methods on simulated data set 2.

	k				
Method	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
RJMCMC	7	0	993	0	0
$\hat{p}_C$	0	12	848	138	2
$\hat{p}_H$	0	0	650	228	122
$\hat{p}_{NR}$	0	0	615	258	127
$\hat{p}_{GD}$	0	0	998	2	0
$\hat{p}_{LM}$	0	1	999	0	0
$\hat{p}_G$	0	11	985	4	0
$\hat{p}_{opt}$	0	11	985	4	0
Criterion	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
AIC	0	0	857	125	18
BIC	0	0	995	5	0
BIC*	0	0	993	7	0
ICOMP	0	0	886	97	17

6.3. Factor structure in international exchange rates

We now explore the factor structure underlying the changes in monthly international exchanges rates using the data studied in West and Harrison (1997, pp.610-618). These time series are the exchange rates in British pounds of the following $m = 6$ currencies: US dollar (US), Canadian dollar (CAN), Japanese yen (JAP), French franc (FRA), Italian lira (ITA) and the (West) German (Deutsch)mark (GER). The data span the period from 1/1975 to 12/1986 inclusive, and the monthly changes in exchange rates appear in Figure 2. Each series has been standardised with respect to its sample mean and standard deviation over the period for comparability (this does not affect the modeling process and factor structure analysis). Earlier studies in West and Harrison (1997) used various principal component analyses that indicated up to three meaningful latent components, suggesting up to three factors may be relevant in our analyses.

Our illustrative analysis first explores uncertainty about the number of factors as in the foregoing simulated data analyses, and then investigates questions about the dependence of conclusions from such analyses on the chosen order of the series. This latter point is of interest as the particular factor model structure adopted − with the upper triangle of zero elements in the factor loading matrix − introduces an apparent order dependence.

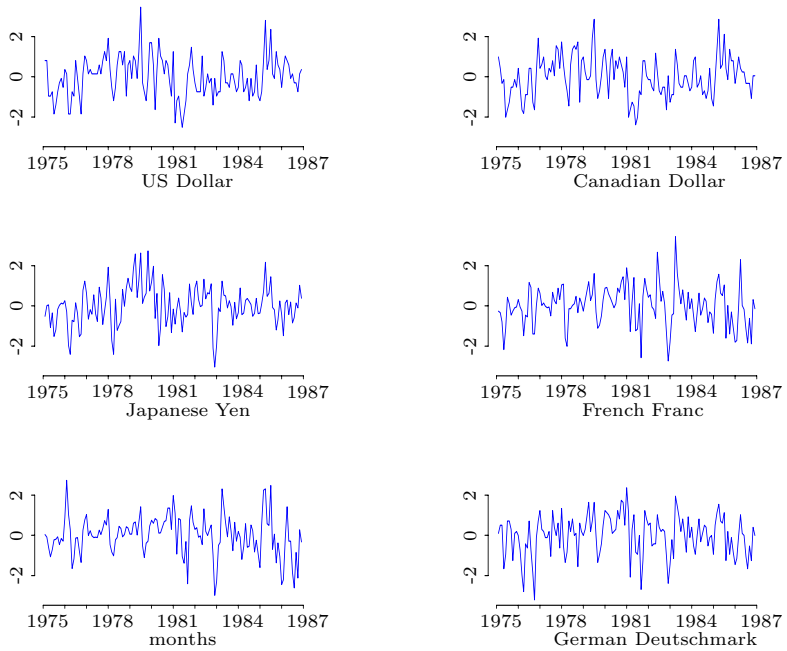

Figure 2. Standardized first differences of monthly observed exchange rates.

Prior distributions are specified exactly as in the simulated examples, the general scales for all parameters now being comparable with those of the simulation examples as the data are modelled after standardisation. Specifically, we have $\mu_0 = 0$, $C_0 = 1$, $\nu_{0i} = 2.2$ and $\nu_{0i}s_{0i}^2 = 0.1$. For the Gibbs sampling and RJMCMC analyses we burn-in the algorithms for $10,000$ iterations, and then save equally spaced samples of $5,000$ draws from a longer run of $100,000$. Newton and Raftery (1994) suggest using δ small, so we decided to use $\delta = 0.05$. Alternative values were used and little or no variation was observed. Proposal distributions in the RJMCMC analysis are based on defining parameters $a = 18$ and $b = 2$, and the transition matrix, J, is such that $J_{ii} = 0$ for $i = 1, 2, 3$ and $J_{ij} = 0.5$ for $i \neq j$.

The analyses were run on the data with currencies ordered by country as: US, CAN, JAP, FRA, ITA and GER. Table 3 provides summaries of the various approximate Bayesian and information criteria for assessment of the number of factors. The overall suggestion is that $k = 2$ is strongly favoured.

From the MCMC analysis of the $k = 2$ factor model, we have the following posterior summaries.

- The posterior means of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ parameters are, to two decimal places,

$$E(\boldsymbol{\beta}'|\boldsymbol{y}) = \begin{pmatrix} 0.99 & 0.95 & 0.46 & 0.39 & 0.41 & 0.40 \\ 0.00 & 0.05 & 0.42 & 0.91 & 0.77 & 0.77 \end{pmatrix}$$

and $E(diag(\boldsymbol{\Sigma})|\boldsymbol{y}) = (0.05, 0.13, 0.62, 0.04, 0.25, 0.28)$.

Table 3. Comparison of model uncertainty assessment methods from analyses of the international exchange rate time series.

	$\log p(\boldsymbol{y}\|k)$			$p(k\|\boldsymbol{y})$		
Method	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
RJMCMC	—	—	—	0.00	**0.88**	0.12
$\hat{p}_C$	-1013.5	-935.3	**-925.5**	0.00	0.00	**1.00**
$\hat{p}_H$	-988.0	**-871.0**	-871.8	0.00	**0.71**	0.29
$\hat{p}_{NR}$	-991.9	**-880.1**	-881.4	0.00	**0.78**	0.22
$\hat{p}_{GD}$	-1017.7	-907.1	**-906.4**	0.00	0.34	**0.66**
$\hat{p}_{LM}$	-1014.8	-904.5	**-897.3**	0.00	0.00	**1.00**
$\hat{p}_G$	-1014.5	-903.7	-Inf	0.00	**1.00**	0.00
$\hat{p}_{opt}$	-1014.5	-903.7	-Inf	0.00	**1.00**	0.00
Criterion	$k=1$	$k=2$	$k=3$			
AIC	1978.4	**1745.0**	1751.0			
BIC	2013.9	**1795.4**	1813.2			
BIC*	2013.6	**1794.8**	1812.3			
ICOMP	1957.9	1776.1	**1724.0**			

- The marginal posterior densities of the elements of $\boldsymbol{\beta}$ are displayed – in terms of histograms of the posterior samples – in Figure 3.
- The marginal posterior densities of the σ_k^2 parameters are displayed – also in terms of histograms of the posterior samples – in Figure 4.
- The time trajectories of the posterior means of the two factor time series are displayed in Figure 5. The first factor is plotted together with the US and CAN series, and the second factor is plotted together with the JAP and European currencies.

Table 4. Percentage of the variance of each series explained by each factor in analysis of the international exchange rate time series.

Country	Factor 1	Factor 2
US	95.1	0
CAN	87.6	0.2
JAP	20.5	17.6
FRA	14.7	81.8
ITA	16.4	58.6
GER	16.1	58.5

- For each currency series $i = 1, \ldots, 6$, the percentage of the conditional variance explained by each factor $j = 1, 2$ is simply $100(1 + \beta_{kk'}^2/\sigma_k^2)$. Table 4 below provides estimated values of these quantities with the $\beta_{\cdot}$ and $\sigma_{\cdot}$ parameters estimated at their posterior means.

These summaries indicate the following broad conclusions. The first factor represents the value of sterling relative to a basket of currencies in which the North American currencies are dominant. US and CAN are roughly equally weighted, which is expected as CAN rates are heavily determined in international markets by US rates. This first factor may be termed the North American factor. The second factor may be similarly termed the European Union (EU) factor. It represents a restricted basket of currencies dominated by the EU currencies, with a relatively reduced weighting on JAP. US is absent from this factor, by design of the factor loading matrix, and CAN is practically absent, with the posterior for $\beta_{2,2}$ indicating very small values. Inferences about the idiosyncratic variances strengthen and extend these conclusions. Those for US and GER are very small, indicating that these two currencies play determining roles in defining their sector factor. CAN, FRA and ITA have larger idiosyncratic variances, indicative of their departures from their sector factors. JAP has a large idiosyncratic variance, contributing about two-thirds of the overall conditional variance. A $k = 3$ factor model would move most of this variability over to the third, JAP factor, as further studies verify.

Notice that the marginal posteriors graphed are all unimodal. This is of interest in view of the earlier discussion about multiple posterior modes induced by multiple local maxima in the likelihood functions when the specified value of k is larger that is consistent with the data. Multiple modes appear in analysis of a 3-factor model very much in the same way as seen in Figure 1. The margins from the 2-factor model analysis are consistent with the view that $k = 2$ is not too large, and therefore provide additional support for the 2-factor model.

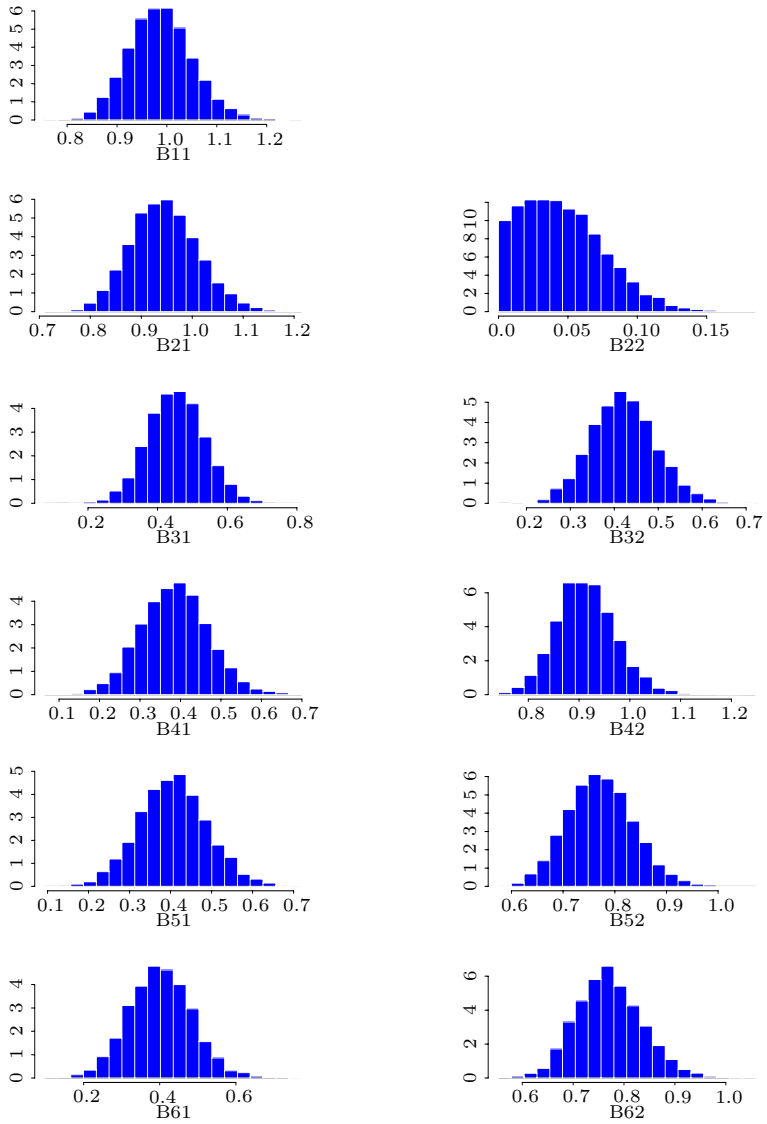

Figure 3. Marginal posteriors of the factor loadings when fitting a two-factor structure to the international exchange rates.

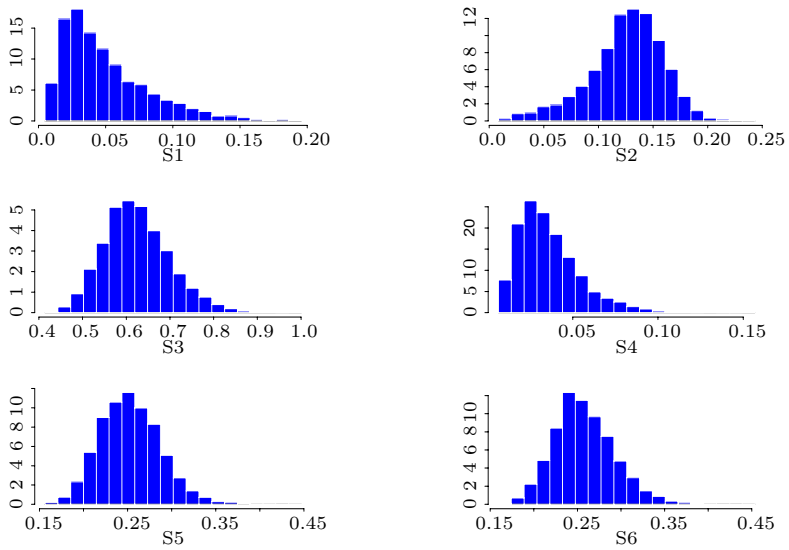

Figure 4. Marginal posteriors of the idiosyncratic variances when fitting a two-factor structure to the international exchange rates.

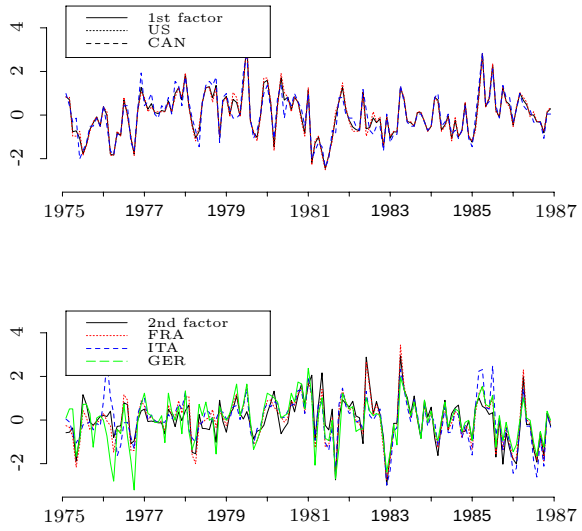

Figure 5. Posterior mean for the factor and actual exchange rates. First factor plus US and CAN (*upper frame*) and second factor plus FRA,ITA and GER (*lower frame*)

It is of interest to explore possible dependence on the order of the series in this analysis. This is especially indicated here due to the high dependence

between US and CAN and the resulting very small values of $\beta_{2,2}$, the diagonal factor loading element that is constrained to be positive. The above analysis was therefore re-run with the orders of CAN and JAP interchanged. The resulting posterior means of the factor loadings and idiosyncratic variances are

$$E(\boldsymbol{\beta}|\boldsymbol{y}) = \begin{pmatrix} 0.98\ \mathbf{0.45}\ \mathbf{0.95}\ 0.39\ 0.41\ 0.41 \\ 0.00\ \mathbf{0.42}\ \mathbf{0.03}\ 0.91\ 0.77\ 0.77 \end{pmatrix}$$

and $E(diag(\boldsymbol{\Sigma})|\boldsymbol{y}) = (0.06, \mathbf{0.62}, \mathbf{0.12}, 0.04, 0.25, 0.26)$, where the figures in bold font simply indicate JAP and CAN in the new order. Comparing with the original analysis we see that these numbers are in extremely close agreement, suggesting the order has essentially no effect on the estimation. To further explore this, Table 5 provides the summaries of model assessment methods based on this ordering of variables. The overall conclusion that $k = 2$ is the "true" number of factors stands.

It is significant that in Table 4, nearly all of the methods pick $k = 2$, as opposed to Table 5. This gives us a hint of how instable some methods are. It might also show that for the less stable methods, the ordering is more important.

Table 5. Comparison of model uncertainty assessment methods from the reanalyses of the international exchange rate time series under a different order of the currencies.

	$\log p(\boldsymbol{y}\|k)$			$p(k\|\boldsymbol{y})$		
Method	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
RJMCMC	—	—	—	0.00	**0.98**	0.02
$\hat{p}_C$	-1013.5	**-934.5**	-985.8	0.00	**1.00**	0.00
$\hat{p}_H$	-988.3	-874.6	**-873.0**	0.00	0.16	**0.84**
$\hat{p}_{NR}$	-985.5	**-867.3**	-867.9	0.00	**0.65**	0.35
$\hat{p}_{GD}$	-1017.8	**-907.0**	-909.6	0.00	**0.93**	0.07
$\hat{p}_{LM}$	-1015.5	**-904.4**	-910.3	0.00	**1.00**	0.00
$\hat{p}_G$	-1014.5	**-903.5**	-Inf	0.00	**1.00**	0.00
$\hat{p}_{opt}$	-1014.5	**-903.5**	-Inf	0.00	**1.00**	0.00
$\hat{p}_{EP}$	-993.5	**-878.4**	-884.2	0.00	**1.00**	0.00

7. Concluding Comments

As discussed in the introduction, our interest has been to study MCMC methods for factor models and novel RJMCMC and other methods for assessing the issue of model uncertainty induced by an unknown number of factors. In doing so, we have explored empirical studies with two simulated and one real data example, highlighting ranges of methodological and modelling issues. A few additional comments are of interest in conclusion.

- Our customised RJMCMC method, using empirical proposal distributions based on parallel MCMC analyses for a range of models with specified numbers of factors, is effective and efficient, as tested in a range of synthetic and real studies (Lopes, Müller and Rosner (2002)). Development of effective proposal distributions and jumping rules in models with even moderate dimensional parameters is usually very challenging, and the approach used here will be useful in other such models. We note that the computation of approximate posterior probabilities on k using bridge sampling approaches, though also requiring some tuning, is similarly accurate in a range of studies, providing answers close to those from RJMCMC; these methods deserve further study too. By contrast, we have found that none of the other "standard" methods of approximation reviewed in Section 5.1 is consistently accurate in identifying correct models in ranges of simulation studies, and none consistently superior to the formal RJMCMC, Newton and Raftery and bridge sampling approaches. On purely empirical grounds, we find the BIC methods generally provide more stable and reliable initial guides to the choice of k than the other standard methods. Traditionally, BIC is an approximation to the log of the Bayes factor for comparing a model to the null model.
- The selected order of data variables influences the likelihood function and hence posterior inferences about the number of factors. The effect can be marked, although inferences about the factor loadings and other parameters are generally relatively stable by comparison. Thus, very naturally, the order of variables is relevant in connection with interpretation of the factors and their number. However, a simple permutation of the variables followed by factors' rotation leads to alternative, and generally more convenient, interpretation of the latent factors, a strategy commonly used by factor analysts. Of course, the conditional variance-covariance matrix of the variables is unaffected by the ordering, and hence so are the predictive inferences resulting from the model.
- In further empirical studies (Lopes (2000)) we have explored a range of predictive exercises. For example, in studies of simulated data and of the industrial stock indices analysed in Geweke and Zhou (1996), we have explored out-of-sample predictions based on sequential data analysis and one-step ahead predictions, as are standard in time series work. These studies have indicated a general point: rather often, posterior uncertainty about the number of factors based on a set of historical data may understate the practically relevant uncertainties when forecasting ahead. For example, sequential forecasting performance using a four-factor model can often out-perform a three-factor model even when analysis of past data has given almost no posterior probability to the four-factor model. Predictive comparisons and model mixing are worthy of further study.

Beyond these issues, we note that related developments in dynamic factor modelling (Aguilar and West (2000), and more recently in Lopes (2000)) in financial time series and portfolio studies are focussed almost exclusively on short-term forecasting and the potential improvements available in forecasting moderate to high-dimensional time series using factor structures. Here the assessment of the number of factors is also a live issue, and one that is complicated by the time-varying nature of such models that leads to the notion of time-variation in the number of (practically relevant) factors.

Acknowledgements

The authors thank Peter Müller of MD Anderson Cancer Center and the referees for insightful comments that further improved the article. This research was partially supported by CAPES, Brazil, and by the National Science Foundation under grants DMS 9704432 and 9707914.

References

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *J. Bus. Econom. Statist.* **18**, 338-357.

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

Arminger, G. and Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271-300.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British J. Math. Statist. Psych.* **34**, 93-99.

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

Besag, J. (1989). A Candidate's formula: a curious result in Bayesian prediction. *Biometrika* **76**, 183-183.

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

Bozdogan, H. and Ramirez, D. E. (1987). An expert model selection approach to determine the "best" pattern structure in factor analysis models. In *Multivariate Statistical Modelling and Data Aanalysis* (Edited by H. Bozdogan and A. K. Gupta).

Bozdogan, H. and Shigemasu, K. (1998). Bayesian factor analysis model and choosing the number of factors using a new informational complexity criterion. Technical Report. Department of Statistics, University of Tennessee.

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

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90**, 1313-1321.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* **68**, 265-274.

Dellaportas, P., Forster, J. J. and Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statist. Comput.* **12**, 27-36.

DiCiccio, T. J., Kass, R. A., Raftery, A. E. and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *J. Amer. Statist. Assoc.* **92**, 903-915.

Gelfand, A. E. and Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *J. Roy. Statist. Soc. Ser. B* **56**, 501-514.

Geweke, J. F. and Singleton, K. J. (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *J. Amer. Statist. Assoc.* **75**, 133-137.

Geweke, J. F. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Rev. Finan. Stud.* **9**, 557-587.

Gilks, R. A., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

Godsill, S. J. (2001). On the relationship between Markov Chain Monte Carlo methods for model uncertainty. *J. Comput. Graph. Statist.* **10**, 230-241.

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

Ihara, M. and Kano, Y. (1995). Identifiability of full, marginal, and conditional factor analysis model. *Statist. Probab. Lett.* **23**, 343-350.

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

Kass, R. A. and Raftery, A. E. (1995). Bayes factor. *J. Amer. Statist. Assoc.* **90**, 773-795.

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

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

Lewis, S. M. and Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *J. Amer. Statist. Assoc.* **92**, 648-655.

Lopes, H. F. (2000). Bayesian analysis in latent factor and longitudinal models. Ph.D. Thesis, Institute of Statistics and Decision Sciences, Duke University.

Lopes, H. F., Müller, P. and Rosner, G. L. (2003). Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics* **59**, 66-75.

Martin, J. L. and McDonald, R. P. (1981). Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases. *Psychometrika* **40**, 505-517.

Meng, X. L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statist. Sinica* **6**, 831-860.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. J. Roy. Statist. Soc. Ser. B **56**, 3-48.

Pitt, M. K. and Shephard, N. (1998). Time varying covariances: a factor stochastic volatility approach. In *Bayesian Statistics* **6** (Edited by J. M. Bernardo et al.) Oxford University Press, London.

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

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

Press, S. J. and Shigemasu, K. (1989). Bayesian inference in factor analysis. *ASA Proc. Social Statistics Section*, 292-294.

Press, S. J. and Shigemasu, K. (1994). Posterior distribution for the number of factors *ASA Proc. Bayesian Statistical Science*, 75-77.

Seber, G. A. F. (1984). *Multivariate Observations*. Wiley, New York.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Soc.* **81**, 82-86.

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

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

Department of Statistical Methods, Institute of Mathematics, Federal University of Rio de Janeiro, Centro de Tecnologia, Bloco C, Caixa Postal 68523, 21945-970, Rio de Janeiro, Brazil.

E-mail: hedibert@im.ufrj.br

Institute of Statistics and Decision Sciences, Box 90251, Duke University, Durham, NC 27708-0251, U.S.A.

E-mail: mw@stat.duke.edu

(Received June 2002; accepted May 2003)