

CAMA

Centre for Applied Macroeconomic Analysis

Invariant Inference and Efficient Computation in the Static Factor Model

CAMA Working Paper 32/2013
June 2013

Joshua C.C. Chan

Centre for Applied Macroeconomic Analysis (CAMA), Australian National University

Roberto Leon-Gonzalez

National Graduate Institute for Policy Studies and
Rimini Center for Economic Analysis

Rodney W. Strachan

Centre for Applied Macroeconomic Analysis (CAMA), Australian National University and
Rimini Center for Economic Analysis

Abstract

Factor models are used in a wide range of areas. Two issues with Bayesian versions of these models are a lack of invariance to ordering of the variables and computational inefficiency. This paper develops invariant and efficient Bayesian methods for estimating static factor models. This approach leads to inference on the number of factors that does not depend upon the ordering of the variables, and we provide arguments to explain this invariance. Beginning from identified parameters which have nonstandard forms, we use parameter expansions to obtain a specification with standard conditional posteriors. We show significant gains in computational efficiency. Identifying restrictions that are commonly employed result in interpretable factors or loadings and, using our approach, these can be imposed ex-post. This allows us to investigate several alternative identifying schemes without the need to respecify and resample the model. We apply our methods to a simple example using a macroeconomic dataset.

Keywords

JEL Classification

Address for correspondence:

(E) cama.admin@anu.edu.au

[The Centre for Applied Macroeconomic Analysis](#) in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Invariant Inference and Efficient Computation in the Static Factor Model*

Joshua Chan

Centre for Applied Macroeconomic Analysis,
Australian National University

Roberto Leon-Gonzalez
National Graduate Institute
for Policy Studies and
Rimini Center for Economic Analysis

Rodney W. Strachan[†]
Centre for Applied Macroeconomic Analysis,
Australian National University and
Rimini Center for Economic Analysis[‡]

June 2013

Abstract

Factor models are used in a wide range of areas. Two issues with Bayesian versions of these models are a lack of invariance to ordering of the variables and computational inefficiency. This paper develops invariant and efficient Bayesian methods for estimating static factor models. This approach leads to inference on the number of factors that does not depend upon the ordering of the variables, and we provide arguments to explain this invariance. Beginning from identified parameters which have nonstandard forms, we use parameter expansions to obtain a specification with standard conditional posteriors. We show significant gains in

*Roberto Leon-Gonzalez and Rodney Strachan are members of the Rimini Center for Economic Analysis.

[†]Corresponding author: Rodney Strachan, Research School of Economics, The Australian National University, 0200 ACT, Australia, email: rodney.strachan@anu.edu.au

[‡]We thank Robert Kohn and Jim Berger for their encouraging comments. We also thank seminar participants at the University of Sydney, the University of Strathclyde, the Multivariate Time Series Modelling and Forecasting Workshop Monash University 2013, the International Workshop on Bayesian Model Selection Shanghai 2013, Workshop, the 18-th Australasian Macroeconomics Workshop, and Workshop on Empirical Methods in Macroeconomic Policy Analysis Bucharest for helpful comments and suggestions. All errors are, of course, our own.

computational efficiency. Identifying restrictions that are commonly employed result in interpretable factors or loadings and, using our approach, these can be imposed ex-post. This allows us to investigate several alternative identifying schemes without the need to respecify and resample the model. We apply our methods to a simple example using a macroeconomic dataset.

1 Introduction

Factor models have proven useful in many areas including psychology, genomics, epidemiology, economics and finance and significant advances in computation using Bayesian approaches (for example, Geweke and Zhou (1996), Aguilar and West (2000) and Chib, Nardari, and Shephard (2006)) have made Bayesian analysis of such models feasible for a range of applications. Two problems that have hampered Bayesian inference in factor models are, first, the models are not invariant to different ordering of the variables (see, for example, Lopes and West (2004)) and, second, poor efficiency of computation algorithms (e.g., Chib *et al.* (2006)).

This paper makes a number of contributions. i) We provide a formal explanation for why extant specifications are not invariant, that is, why the evidence in the model can change when the order of the variables changes. ii) This paper presents an invariant specification. That is, the specification will result in inference that does not depend upon the ordering of the variables. iii) We use parameter expansions to develop an algorithm that is both easy to implement and computationally efficient. The resulting posteriors have relatively simple normal forms. iv) We demonstrate that there is not an identification problem so much as a specification problem in these models. v) We show that some popular Bayesian models are overparameterized. This is not a criticism, as we use parameter expansions to overparameterize our model also. However, we follow the rules of Liu and Wu (1999) to ensure efficiency gains. It is not clear that extant specifications do follow these rules and this may explain to some degree the poor sampling.

Our proposed approach to obtaining invariant inference on the number of factors and efficient computation uses a specification which is commonly used in the frequentist literature (see, for example, Connor and Korajczyk (1986) and Bai and Ng (2002)) combined with an extension of the parameter expansion approach of Ghosh and Dunson (2009) in the static factor model and Koop, León-González and Strachan (2010 & 2013) in the vector error correction model and instrumental variables model.

Reordering of variables involves groups of transformations of the parameters in the model. We therefore use group theory to show why existing specifications are not invariant to reordering of the variables. Work to date considering invariance has taken one of two approaches. The first approach attempts to resolve the issue by averaging over orderings (see for example Geweke (1996) and Frühwirth-Schnatter and Lopes (2010)). To estimate k we would need to esti-

mate all orderings for all values of k . Averaging over orderings shows promise in small dimensional settings, but as applications often have many variables (sometimes hundreds), the number of potential orderings to average over increases into the trillions making an averaging approach computationally infeasible or at best challenging. For example, Forni, Giannone, Lippi and Reichlin (2009) investigate some 89 series and find there are between $k = 12$ and $k = 18$ factors. In this case, if we were to use a non-invariant Bayesian approach and average over all orderings we would need to average over more than 237 trillion for $k = 12$ and 3 million trillion models for $k = 18$. It would seem more practical and feasible to only have one invariant model to consider for each k .

Another approach, therefore, is to develop a single model that does not depend upon the orderings. Examples of work taking this approach, besides our paper, are Aßmann, Boysen-Hogrefe, and Pape (2012) and Kaufmann and Schumacher (2012). Our approach differs from that of Aßmann, Boysen-Hogrefe, and Pape (2012) and Kaufmann and Schumacher (2012) in that we explicitly take the perspective of the factor model as a reduced rank regression model, such as in Bai and Ng (2002), and use previous work that utilizes the geometry of that model to develop an invariant model specification and inferential framework. Taking this perspective leads us to the view that, contrary to general belief, there is not an identification problem in the factor model but rather there is a problem with the specification used.

As the invariant specification we develop has a structure that would be difficult to compute, we use parameter expansions to develop a specification that is simple to implement, in fact simpler than standard extant specifications. A further benefit of this expanded specification is that the resulting sampler is efficient. This parameter expansion may be viewed as a generalization of the Ghosh and Dunson (2009) approach to computing factor models but with the added benefit of invariance.

In Section Two we briefly outline the features of the static factor model and discuss relationships, in particular mappings, among existing identification schemes, two of which are popular non-invariant specifications used in the Bayesian literature and one invariant specification used in the frequentist literature. In Section Three we present the identified parameters in the invariant Bayesian specification of the static factor model and the priors for this model from a singular value decomposition (SVD). We then introduce the full parameter expansion using invariant transformations to obtain the prior for the ‘expanded’ model. We present the posterior, sampling algorithm and the posterior probability estimation for this ‘expanded’ model. Section Four presents a small application to six exchange rates to demonstrate the effect of reordering and efficiency of the proposed sampling algorithm in this paper. Section Six provides some concluding comments and potential extensions.

2 Alternative specifications of the static factor model:

In this section we outline some standard identification schemes that have been used in static and sometimes dynamic factor models. This section provides an incomplete overview of the literature but focusses the discussion on relevant models. We use two Bayesian schemes to demonstrate issues of invariance. A third approach, a frequentist approach, is an invariant specification but not easily computable using Bayesian techniques in its current form.

A change in the order of the variables involves a transformation of the loading matrix and this transformation involves a discontinuity. We demonstrate the importance of this discontinuity for inference on the number of factors with a simulation. We present the factor model developed using principle components as in Bai and Ng (2002) and show how these could be estimated using Bayesian methods. This gives a feel for the approach developed in the next section. Finally, we show how to map from the principal components specification to the other model specifications discussed in this section and this demonstration also suggests how our specification can permit any of the others.

Noninvariant specifications: Denote the $1 \times n$ vector of observables of variables of interest by y_t . A simple static factor model for y_t with $k < n$ factors is

$$y_t = f_t \Lambda + \varepsilon_t \quad (1)$$

$$\begin{pmatrix} \varepsilon_t' \\ f_t' \end{pmatrix} \sim N\left(0, \begin{bmatrix} \Sigma & 0 \\ 0 & \Omega \end{bmatrix}\right) \quad (2)$$

where f_t is a $1 \times k$ vector, Λ is an $k \times n$ matrix, $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ and for now we simply assume Ω is a general $k \times k$ symmetric positive definite matrix. We can readily add exogenous variables to this specification however we ignore these here for simplicity and to focus upon the specification issues. The variance of y_t is

$$E(y_t' y_t) = \Lambda' \Omega \Lambda + \Sigma.$$

The identification issues with this model are well understood. The unrestricted Λ , Ω and Σ contain a total of $nk + n + k(k+1)/2$ parameters which can exceed the $n(n+1)/2$ parameters in $E(y_t' y_t)$. Even if we assume $\Omega = I_k$, we must restrict $k \leq (n-1)/2$ as otherwise there would not be enough information to estimate all of the parameters (Geweke and Zhou (1996)). There remains a second identification problem as we can rotate f_t and obtain an observationally equivalent vector f_t^* . That is, if $U \in O(k) \equiv \{U : U(k \times k), U'U = I_k\}$ the orthogonal group, then rotate the parameters by $f_t^* = f_t U$ and $\Lambda^* = U' \Lambda$ and we can see that the rotated parameters (f_t^*, Λ^*) and (f_t, Λ) are not distinguishable since they both enter the likelihood as products: $f_t^* \Lambda^* = f_t \Lambda$. For this reason restrictions are imposed upon Λ and Ω to permit estimation. Note that this argument uses a specification of f_t and Λ that is overidentified and this permits the observationally equivalent rotations. For the following discussion it

is useful to denote the first k columns of Λ by Λ_1 . Generally we will take the matrix Λ_1 to represent the columns that are restricted to permit identification of the factors and these may not always be the first k columns.

An identification scheme used, for example, in Geweke and Zhou (1996) sets $\Omega = I$ and restricts $\Lambda = \Lambda^+ \in \mathfrak{T}^+$ where \mathfrak{T}^+ is the space of upper triangular matrices with positive elements on the diagonal. We will also use the notation \mathfrak{T}^+ for the identification scheme in this approach. Thus the specification is

$$\begin{aligned}
 y_t &= f_t^+ \Lambda^+ + \varepsilon_t \\
 \begin{pmatrix} \varepsilon_t' \\ f_t^{+'} \end{pmatrix} &\sim N\left(0, \begin{bmatrix} \Sigma & 0 \\ 0 & I \end{bmatrix}\right) \\
 \Lambda^+ &= \begin{bmatrix} + & * & \cdots & * & * & \cdots & * \\ 0 & + & & * & * & \cdots & * \\ \vdots & & \ddots & * & * & \cdots & * \\ 0 & 0 & & + & * & & * \end{bmatrix} \\
 &= [\Lambda_1^+ \quad \Lambda_2]
 \end{aligned} \tag{3}$$

where the plus signs denote parameters restricted to be positive and the asterisks denote unrestricted elements. The first factor is identified as it is the only factor entering the first variable. The next factor is identified as, besides the already identified first factor, it is the only factor entering the second variable, and so on. This structure requires the researcher to assume some knowledge about the order of the variables with respect to the factors which imply the first k rows of the loading matrix are linearly independent.

If we collect the diagonal elements of Λ_1^+ from the \mathfrak{T}^+ specification into the diagonal of an $k \times k$ matrix Λ_d^+ , we can transform to another specification that is often used. Define $f_t^1 = f_t^+ \Lambda_d^+$ and $\Lambda^1 = (\Lambda_d^+)^{-1} \Lambda^+$ and $\Lambda_1^1 = (\Lambda_d^+)^{-1} \Lambda_1^+$ such that $f_t^+ \Lambda^+ = f_t^+ \Lambda_d \Lambda_d^{-1} \Lambda^+ = f_t^1 \Lambda^1$ and let $\Omega^1 = \Lambda_d^{+'} \Lambda_d^+$ be a diagonal covariance matrix for the errors in the factor process. In this identification scheme, $\Lambda = \Lambda^1 \in \mathfrak{T}^1$ where \mathfrak{T}^1 is the space of upper triangular matrices with ones on the diagonal. This gives us the second specification,

$$\begin{aligned}
 y_t &= f_t^1 \Lambda^1 + \varepsilon_t \\
 \begin{pmatrix} \varepsilon_t' \\ f_t^{1'} \end{pmatrix} &\sim N\left(0, \begin{bmatrix} \Sigma & 0 \\ 0 & \Omega^1 \end{bmatrix}\right) \\
 \Lambda^1 &= \begin{bmatrix} 1 & * & \cdots & * & * & \cdots & * \\ 0 & 1 & & * & * & \cdots & * \\ \vdots & & \ddots & * & * & \cdots & * \\ 0 & 0 & & 1 & * & & * \end{bmatrix} \\
 &= [\Lambda_1^1 \quad \Lambda_2^1]
 \end{aligned}$$

This specification has been used in, for example, Chib, Nardari and Shephard (2006) and Aguilar and West (2000) and requires similar assumptions about the orders of the variables as the previous specification.

Notice that these specifications have the same dimension of parameter space as Σ is the same, f_t is always $1 \times k$, and the free elements of $[\Lambda^+, \Omega^+]$ and $[\Lambda^1, \Omega^1]$ all have dimension $(n - k)k + \frac{k(k+1)}{2} = nk - \frac{k(k-1)}{2}$. With all f_t , the total number of free parameters is then $(T + n)k - \frac{k(k-1)}{2}$. These non-invariant identification schemes (and priors) used in Bayesian approaches result in specifications that have more parameters than can be identified from the likelihood. Thus they are still not identified under the likelihood, but are identified under the posterior due to the informative priors on the overidentified parameters.

The source of the lack of invariance: A drawback of these two specifications is that they assume Λ_1 is nonsingular. In particular, they assume that the first k variables in y_t are functions of all of the first k factors. As many practitioners are aware, these assumptions have implications for evidence on alternative values for k . Transforming the parameters by reordering the variables involves groups of transformations with a discontinuity at the point $|\Lambda_1| = 0$. This discontinuity implies the transformation is non-homeomorphic and results in the non-invariant and so non-unique inference. We provide a small simulation study to provide a further illustration of the issue and to demonstrate that this effect is neither ubiquitous nor a measure zero event. That is, the effect does not occur at all points of the parameter space but nor is it only relevant at the point $|\Lambda_1| = 0$.

For our simulation study, let the model be as before but add an intercept,

$$y_t = \mu + f_t \Lambda + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma)$$

and let $n = 4$, $T = 600$, $k = 1$ and use the following parameter values:

$$\begin{aligned} \mu &= \{\mu_i\} = (0.1 \quad 1.8 \quad 2.5 \quad 2.5)' \\ \Lambda &= \{\lambda_i\} = (c \quad 0.65 \quad 0.20 \quad 0.20) \\ c &\in \{0, 0.025, 0.05, 0.075, 0.10\} \\ i &= 1, 2, 3, 4 \text{ and } \Sigma = \sigma^2 I_4. \end{aligned}$$

We assume the following set of priors:

$$\begin{aligned} \lambda_1 &\sim N(0, 10) 1(\lambda_1 > 0,) \\ \lambda_i &\sim N(0, 10) \quad i = 2, 3, 4 \\ \mu_i &\sim N(0, 10) \quad i = 1, 2, 3, 4 \\ \sigma^2 &\sim IG(3, 1) \quad i = 1, 2, 3, 4. \end{aligned}$$

For various values of c and σ^2 , we compute the marginal likelihoods for two models: a single-factor model with the ordering $(1, 2, 3, 4)$ denoted as $(1, 2)$, so that c is the loading for the first variable in the DGP λ_1 , and another single-factor model with the ordering $(2, 1, 3, 4)$ denoted as $(2, 1)$, so that c is the loading for the second variable in the DGP λ_2 . The ordering $(1, 2)$ is incompatible with the identification scheme when $c = 0$. The marginal likelihoods are computed via the Chib's method using 20 parallel chains each of length 10000. The results are reported in Table 1:

Table 1: Bayes factors (relative numerical standard errors) for the models under different orderings

c	$BF_{(1,2),(2,1)} (NSE_{(1,2)}/NSE_{(2,1)})$		
	$\sigma^2 = 0.05$	$\sigma^2 = 0.10$	$\sigma^2 = 0.20$
0.00	11.5 (1.67)	10.07 (45.71)	7.39 (32.86)
0.025	1.32 (1.67)	2.51 (6.67)	6.69 (11.25)
0.050	1.01 (0.75)	1.31 (13.33)	2.46 (1.50)
0.075	1.02 (1.00)	1.01 (1.00)	1.49 (1.00)
0.100	1.00 (0.75)	0.99 (1.00)	1.11 (0.75)

As the marginal likelihood for the model with no factors is the same for both orderings, we report in Table 1 the ratio of the marginal likelihoods for the model under the ordering (2, 1, 3) to the model under the ordering (1, 2, 3). This is just the Bayes factor for the models under the different orderings. If reordering had no effect, that is, if the models were invariant to ordering, the Bayes factors would all be one. Any difference from one is due to the lack of invariance. Clearly around the point $c = 0$ where the ordering (1, 2, 3) is incompatible with the identification scheme, we see large differences of between 7 and 11 in the marginal likelihoods. As c increases the effect disappears although less quickly for the larger σ^2 . It appears that both of the parameters c and σ^2 each play a role in defining a metric for distance from the point $c = 0$.

The results in these tables give an indication of how lack of invariance to the ordering of the variables manifests in the evidence on k . Using real data we tend to find the evidence is much stronger (see, for example, Lopes and West (2004) and the empirical application in this paper).

Lack of invariance has an interesting and we think important effect upon the quality of inference. In the parentheses in Table 1 we report the ratio of the numerical standard errors for the marginal likelihoods computed under the different orderings, $(NSE_{(1,2)}/NSE_{(2,1)})$. This ratio shows the second effect of non-invariance; the poorer quality of the estimator of the marginal likelihood. At low values of c the numerical standard errors for the model (1, 2) are much larger for large σ^2 than for the models (2, 1). The relative efficiency of the estimator improves as we move away from $c = 0$. Clearly if we are relying upon marginal likelihoods for inference, we should have less confidence in our conclusions in this case.

Given the above evidence, it would sensible to choose the ordering in which the posterior mass for $|\Lambda_1|$ is far from zero. Unfortunately, this point cannot be known a priori which brings us back to the solutions of either averaging over all orderings or devising an invariant specification.

A common representation of the factor model is as a reduced rank regression model. In reduced rank models the row and column space of the reduced rank matrix are identified under the likelihood (Villani, 2005). The problem of invariance in the factor models stems from the transformation from these spaces to the parameters in the model to be estimated, and from the transformation between different models that can be estimated. These transformations do not

preserve the topology such that changing the specification used, or the ordering, can result in a change in the evidence on the number of factors. To introduce notation, denote the space of an $m \times k$ matrix H by $sp(H)$. The support for $sp(H)$ is the collection of all k -planes in m -space. This collection is called the Grassmann manifold which we denote by $G_{k,m-k}$ such that $sp(H) \in G_{k,m-k}$. $G_{k,m-k}$ is a compact subspace for which there exists a uniform Haar invariant (to left and right orthonormal rotations) measure. The row and column space, $sp(F) \in G_{k,T-k}$ and $sp(\Lambda') \in G_{k,n-k}$, are identified.

The identification schemes \mathfrak{T}^+ and \mathfrak{T}^1 impose an order on the variables.

Definition 1 Denote by $\mathfrak{T}^{m,i}$ ($m \in \{+,1\}$) ordering i of the variables and denote a different ordering j by $\mathfrak{T}^{m,j}$.

Definition 2 In the ordering $\mathfrak{T}^{m,i}$ the identification of the factors is achieved by placing restrictions on the submatrix Λ_1^i .

Theorem 3 (Discontinuity) The transformation from $\mathfrak{T}^{m,i}$ to $\mathfrak{T}^{m,j}$ has a discontinuity at $|\Lambda_1^j| = 0$.

Proof. See the Appendix I. ■

To demonstrate this discontinuity with a simple example, consider a (2×1) vector $\Lambda = (\cos(\theta), \sin(\theta)) = (\lambda_1, \lambda_2)$ where $\theta \in [0, \pi)$. The specification \mathfrak{T}^1 implies $\Lambda = (1, \lambda_2)$ where $\lambda_2 = \tan(\theta)$ under one ordering, $\mathfrak{T}^{m,1}$, with support $(-\infty, \infty)$ which includes $\lambda_2 = 0$. Changing the order we obtain $\Lambda = (\lambda_1, 1)$ where $\lambda_1 = \frac{1}{\tan(\theta)} = \frac{1}{\lambda_2}$ is ordering $\mathfrak{T}^{m,2}$ which excludes the point $\lambda_2 = 0$. In this case there is a discontinuity from $\mathfrak{T}^{m,1}$ to $\mathfrak{T}^{m,2}$. As the support of θ has a boundary whereas that of λ_2 does not, $[0, \frac{\pi}{2}) \cap (\frac{\pi}{2}, \pi]$ as it excludes the point $\theta = \frac{\pi}{2}$ and so there is a discontinuity from \mathfrak{T}^1 to $G_{1,1}$. If, however, we used the specification $\Lambda = (\cos(\theta), \sin(\theta)) = (\lambda_1, \lambda_2)$ where $\theta \in [0, \pi)$, then this simple case the support of the angle θ maps homeomorphically to the one dimensional Grassmann manifold, $G_{1,1}$, and changing the order of the variables maps from $G_{1,1}$ to $G_{1,1}$ without discontinuities.

Transforming from $sp(\Lambda')$ to Λ^+ and to Λ^1 does not preserve the topology of these spaces as the deformation ‘punches holes’ in the form at points where $|\Lambda_1| = 0$ as the transformation from the space to Λ^+ is discontinuous at this point. Similarly, transforming between different orderings, changing between Λ_1 and $\tilde{\Lambda}_1$, involves discontinuities at $|\Lambda_1| = 0$. Without preservation of the topology, the metric and therefore measure are not preserved hence the evidence on k under alternative normalizations need not be the same. This effect is naturally most pronounced when the true parameter values are near the point of discontinuity. As our small simulation exercise showed, if we move far enough away from this point the effect is mitigated.

An invariant specification: Next we discuss a third specification that used in the method of principal components (*pc*) to estimate the factors as in, for example, Bai and Ng (2002) (hereafter BN). Denote this model by

$$y_t = f_t^{pc} \Lambda^{pc} + \varepsilon_t.$$

Stacking the observations over T we obtain

$$y = F^{pc} \Lambda^{pc} + \varepsilon$$

where $y = (y'_1, y'_2, \dots, y'_T)'$, $\varepsilon = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_T)'$ and $F^{pc} = (f_1^{pc'}, f_2^{pc'}, \dots, f_T^{pc'})'$.

In BN, the matrix F^{pc} is estimated as proportional to the matrix of eigenvectors¹ associated with the k largest eigenvalues of the matrix yy' . In other words, they take a SVD of y as

$$y = USV' = U_1 S_1 V_1' + U_2 S_2 V_2'$$

where $U = [U_1 \ U_2] \in O(T)$ are the eigenvectors of $yy' = US^2U'$ and $V = [V_1 \ V_2] \in O(n)$ are the eigenvectors of $y'y = VS^2V'$ and $U'U = V'V = I_n$. Setting the factors and loading matrices to $F^{pc} = U_1$ where and $\Lambda^{pc} = S_1 V_1'$, then Λ^{pc} has the unusual property of being orthogonal (but not orthonormal) as $\Lambda^{pc} \Lambda^{pc'} = S_1^2$. The orthonormal structure of F^{pc} and orthogonal structure of Λ^{pc} imply that the parameters are identified up to sign. That is, $F^{pc} \Lambda^{pc} = F^{pc} \kappa \kappa \Lambda^{pc} = F^{pc,*} \Lambda^{pc,*}$ where F^{pc} and $F^{pc,*}$, and Λ^{pc} and $\Lambda^{pc,*}$ will have the same structure only for a diagonal matrix κ with the i^{th} diagonal element equal to $+1$ or -1 .² This lack of identification is resolved by fixing the sign of, say, the first row of U to be non-negative (but not simply positive). Such a restriction gives a particular orientation of the vectors in U in their space, but in no way restricts the space they span and, as the transformation involving κ is homeomorphic, we can regard U, S and V as identified for any practical purposes and, in particular, our purpose. As there exists a specification that is both identified and invariant to ordering, this raises the question as to whether there is an identification problem. Rather, it appears there is a specification problem that induces a lack of invariance.

We mention in passing that there is a difference in the error structure between the estimation of the factors by SVD and our proposed stochastic model for y . The estimation method by principal components is referred to as approximate because the errors are weakly correlated. We can characterize this weak correlation in the following way. As y is a full rank matrix, estimating the matrix $F^{pc} \Lambda^{pc}$ by $U_1 S_1 V_1'$ then implies ε represents $\tilde{\varepsilon} + \nu$ where $\tilde{\varepsilon} = U_2 S_2 V_2'$ and ν is a matrix of independent errors (i.e., $E(\text{vec}(\nu) \text{vec}(\tilde{\varepsilon})') = 0$ and $E(\text{vec}(\nu) \text{vec}(\nu)') = I_T \otimes D$ where D is diagonal positive definite). The $T \times n$ matrix $\tilde{\varepsilon}$ is rank $(n - k)$ while in our model the matrix ε is full rank. This distinction does not seem to be of major importance.

This model is usually estimated using a nonparametric approach in that no distribution is assumed for f_t^{pc} , and sometimes no distribution is assumed for ε_t . We are not aware of any attempt to estimate this specification using a Bayesian approach but we discuss this possibility here as it most closely resembles the specification we begin with prior to parameter expansion.

¹In fact, Bai and Ng use \sqrt{T} times the eigenvalue of yy' . This proportional term is not important for the discussion here so we ignore it.

²If κ were not of this structure then the restriction $\Lambda^{pc} \Lambda^{pc'} = S_1^2$ would be destroyed by the transformation. Thus this restriction implies identification against such a transformation.

The structures of F^{pc} and Λ^{pc} are difficult to implement in a Bayesian computational approach. To permit Bayesian estimation one could introduce a parameter expansion. That is, an unidentified orthonormal parameter, C such that $C'C = I_k$, is introduced and $F^{*pc} = UC'$ and $\Lambda^{*pc} = CSV$ such that $\Lambda^{*pc}\Lambda^{*pc'} = CS_1^2C'$ is now a full real matrix. We will use a similar idea in developing our model for computation.

The restriction $F^{*pc'}F^{*pc} = I_k$ implies that the support for the matrix of factors, F^{*pc} , is the Stiefel manifold which we denote as $V_{m,n} = \{H (n \times m); H'H = I_m\}$ and so $F^{*pc} \in V_{k,T}$ (for discussion, see Muirhead (1982)).

A purpose of this section is to show how we can move from one non-invariant specification to another via simple transformations. The existence of this mapping is important as we will show how we can map from our invariant specification to one of those in this section, which implies we can then map to both specifications from ours. The transformation from one non-invariant specification to another, with a given ordering of the variables, is possible because all of the specifications impose the same restriction on the parameter space. That is, they all assume that a particular set of k columns are linearly independent. This assumption does not hold at all points in the parameters space when changing the order of the variables so it is not possible to map homeomorphically between specifications with different variable orderings.

It is possible to map from (F^{pc}, Λ^{pc}) to either \mathfrak{T}^+ or \mathfrak{T}^1 with parameter expansion. We show here the map from (F^{pc}, Λ^{pc}) to either \mathfrak{T}^+ . First, introduce an $k \times k$ rank k matrix with $\frac{k(k+1)}{2}$ free parameters κ via

$$\begin{aligned} U_1 S_1 V_1' &= U_1 \kappa \kappa^{-1} S_1 V_1' = F \Lambda \\ \text{where } F &= U_1 \kappa \text{ and } \Lambda = \kappa^{-1} S_1 V_1'. \end{aligned}$$

Next take a QR factorization (see, for example, Schott (1997)) of Λ

$$\Lambda = C \Lambda^+ \text{ where } C'C = I_r \text{ and } \Lambda^+ \in T^+$$

and define

$$\begin{aligned} F \Lambda &= F^+ \Lambda^+ \\ \text{where } F^+ &= FC. \end{aligned}$$

There are $(T + N)k - k^2$ identified parameters in U_1, S_1 and V_1 while there are $(T + N)k - \frac{k(k-1)}{2}$ parameters in F^+ and Λ^+ . The above transformation using the $\frac{k(k+1)}{2}$ extra ‘expanding’ parameters in κ show how \mathfrak{T}^+ is an over-parameterized specification. Liu and Wu (1999) show that improvements in efficiency are achieved if the priors for the identified parameters and the expanding parameters are independent, and if the transformations involved are diffeomorphisms. These conditions do not seem to be met in the transformation $(U_1, S_1, V_1) \rightarrow (F^+, \Lambda^+)$. The QR factorization above requires the first k columns of Λ be of full rank which implies a discontinuity at points where this does not hold which suggests a source of computational inefficiency. The Jacobian for the transformation $(U_1, S_1, V_1, \kappa) \rightarrow (F^+, \Lambda^+)$ has a somewhat complex

form and inverting it is not straightforward. At this point we have not been able to use this and the usual priors on (F^+, Λ^+) to determine whether the implied priors on κ and (U_1, S_1, V_1) are independent.

In the following section we provide the technical details for the several contributions of this paper. The reader who prefers not to read the technical details in Section 3 and is interested only in applying the approach may prefer to skip to Section 4.

3 The invariant static factor model

In this section we develop the prior for the expanded model from independent priors for the factors and the loading matrix. We begin with a statement of what parameters are identified under the likelihood and show how there is not, in fact, an identification problem in this model but there is a specification problem. By this we mean that the parameters of the factor model are all identified. They are not, however, readily computable as they have very non-standard structures. Diffeomorphic transformations to standard structures permit invariant (order independent) inference on the number of factors.

To this point we have said little about the priors for Λ and have only considered distributions for f_t that have previously been used in the literature. In this section we develop priors for Λ and f_t by beginning with the parameters that are identified under the likelihood and, via a series of parameter expansions, we obtain the prior and posterior for the expanded model.

Stacking the y_t , f_t and ε_t into vectors and matrices y , F and ε we can write the factor model as a reduced rank regression model

$$y = F\Lambda + \varepsilon = \Pi + \varepsilon \tag{4}$$

where the matrix $\Pi = F\Lambda$ has rank k . The reduced rank model in (4) has the same structure as a one-mode analysis used in psychometrics (see, for example, Magnus and Neudecker (1988)) for which frequentist approaches to estimation are proposed. Bayesian inference in other reduced rank models, such as the cointegrating vector error correction model and the overidentified simultaneous equations model, has been extensively explored and this literature is informative on how to approach the analysis of this model. Bayesian approaches most relevant to this paper are Strachan and Inder (2004), Koop, León-González and Strachan (2010 & 2013).

To avoid such discontinuous transformations, in the next section we present a specification that involves homeomorphic transformations, which permits invariant inference on the number of factors, and parameter expansions that leads to simple and efficient sampling schemes. Taking a singular value decomposition (SVD) of the reduced rank matrix $F\Lambda$, we have

$$\begin{aligned} F\Lambda &= USV' \\ U_1 &\in V_{k,T} \quad V_1 \in V_{k,n} \\ S_1 &= \text{diag}(s_1, s_2, \dots, s_k) \quad \text{where } s_i > 0 \quad \forall_i. \end{aligned}$$

All of the parameters U_1 , S_1 and V_1 are identified and have, respectively, $Tk - \frac{k(k+1)}{2}$, k and $nk - \frac{k(k+1)}{2}$ free elements.

We specify independent priors for U_1 , S_1 and V_1 with the form

$$f(S_1) \frac{(dS_1)(U_1' dU_1)(V_1' dV_1) c_O}{c_N c_U}$$

$$f(S_1) \propto \exp\left\{-\frac{c_\lambda}{2} \text{tr} S^2\right\} 2^{-k} |S|^{n-k} \prod_{i < j}^k (s_i^2 - s_j^2),$$

$$\frac{c_N}{c_O} = \int f(S_1) (dS_1) (V_1' dV_1),$$

$$c_U = \int (U_1' dU_1) = \frac{2^k \pi^{\frac{Tk}{2}}}{\Gamma_k\left(\frac{T}{2}\right)},$$

$$\Gamma_k\left(\frac{T}{2}\right) = \pi^{k(k-1)/2} \prod_{i=1}^k \Gamma\left[\frac{T-i+1}{2}\right],$$

where $U_1 \in V_{r,T}$ and $V_1 \in V_{r,N}$ have uniform priors on Stiefel manifolds (for further discussion see James (1954)). We give explicit expressions for c_N and c_O below. The diagonal elements of the matrix S_1 have a ‘standard’ prior which we will show is implied by a normal for $\text{vec}(\Lambda) \sim N\left(0, \frac{1}{c_\lambda} I_{Nr}\right)$. The priors for U_1 , S_1 and V_1 are all proper. The term c_λ is included to permit shrinkage of Λ towards zero or a more diffuse prior if desired.

We introduce the nonidentified parameters by two parameter expansions. From the first expansion, we obtain a normal form for the loading matrix and the second expansion results in a normal prior for the factors. These expansions do not affect the proper, independent priors for U , S and V , uniform priors for U and V , and the standard prior for S .

Map from the SVD parameters (U_1, S_1, V_1) to the expanded parameters (U_1, Λ^*) by introducing the orthogonal matrix C via the transformation

$$U_1 S_1 V_1' = U_1 C' C S_1 V_1' = U \Lambda^*$$

$$U_1 C' = U \text{ and } C S_1 V_1' = \Lambda^*$$

$$\text{where } C' C = I_r \text{ since } C \in O(r).$$

The expanding parameter C is given a uniform distribution on $O(r) : (C' dC)$. Using results in, for example, Muirhead (1982), the prior now becomes

$$f(S_1) \frac{(dS_1)(U_1' dU_1)(V_1' dV_1)(C' dC)}{c_N c_U} = p^*(\Lambda^*) \frac{(d\Lambda^*)(U' dU)}{c_N c_U}$$

$$\text{where } p^*(\Lambda^*) = \exp\left\{-\frac{c_\lambda}{2} \text{tr} \Lambda^* \Lambda^*\right\},$$

$$\begin{aligned}
c_N &= \int p^*(\Lambda^*) (d\Lambda^*) \\
&= \int f(S_1) (dS_1) (V_1' dV_1) (C' dC) \\
&= \left(\frac{2\pi}{c_\lambda} \right)^{\frac{nk}{2}}, \\
c_U &= \int (U_1' dU_1) = \int (U' dU) = \frac{2^k \pi^{\frac{Tk}{2}}}{\Gamma_k\left(\frac{T}{2}\right)}, \\
c_O &= \int (C' dC) = \frac{2^k \pi^{\frac{k^2}{2}}}{\Gamma_k\left(\frac{k}{2}\right)}, \\
\Gamma_k\left(\frac{m}{2}\right) &= \pi^{k(k-1)/2} \prod_{i=1}^k \Gamma\left[\frac{m-i+1}{2}\right],
\end{aligned}$$

and so $\Lambda^* = CS_1V_1'$ has a normal prior distribution such that $p^*(\Lambda^*)$ has a form proportional to the density of a zero mean normal distribution with covariance matrix $\frac{1}{c_\lambda} I_{Nr}$. The semi-orthogonal matrix U has a uniform distribution over $V_{r,T}$. In the new parameterization, $F\Lambda = U\Lambda^*$, the matrix Λ^* is has a 'nice' form and prior but U has a very non-standard form and so it would be difficult to obtain draws.

To give the parameters a more standard form and prior distributions, we transform from (U, Λ^*) to (F, Λ) via the second parameter expansion. Introduce the $k \times k$ rank k matrix κ with $\frac{k(k+1)}{2}$ free parameters. κ maybe, for example, lower triangular or symmetric. This matrix is used to obtain the following transformations:

$$U\Lambda^* = U\kappa\kappa^{-1}\Lambda^* = F\Lambda$$

where $F = U\kappa$ and $\Lambda = \kappa^{-1}\Lambda^*$. It is easier to work with the transformation $A = \kappa\kappa' = F'F$ and write the Jacobian as

$$\begin{aligned}
p^*(\Lambda^*) \frac{(dA)(U'dU)(d\Lambda^*)}{c_N c_U} &= p(\Lambda, F) J(F) \frac{(d\Lambda)(dF)}{c_N c_U} \\
J(F) &= 2^k |F'F|^{-(T-n-k-1)/2} \\
p(\Lambda, F) &= \exp\left\{-\frac{c_\lambda}{2} \text{tr} \Lambda' F' F \Lambda\right\}
\end{aligned}$$

Clearly the presence of the determinant $|F'F|$ in the above Jacobian would complicate computation, particularly as we prefer to have a standard form such as a normal distribution for F . Fortunately, we are free to choose the distribution of A and so we let this matrix have a Wishart Distribution with degrees of freedom such that the prior for A is proportional to

$$\exp\left\{-\frac{1}{2} \text{tr} A\right\} |A|^{(T-n-k-1)/2} = \exp\left\{-\frac{1}{2} \text{tr} F'F\right\} |F'F|^{(T-n-k-1)/2}.$$

When we introduce this into the full prior we obtain the following transformation of the measure

$$\begin{aligned}
& |A|^{(T-n-k-1)/2} \exp\left\{-\frac{1}{2}\text{tr}A\right\} p^*(\Lambda^*) \frac{(dA)(U'dU)(d\Lambda^*)}{c_N c_U c_A} \\
= & \exp\left\{-\frac{1}{2}\text{tr}F'F\right\} |F'F|^{(T-n-k-1)/2} p(\Lambda, F) \\
& \times 2^k |F'F|^{-(T-n-k-1)/2} \frac{(d\Lambda)(dF)}{c_N c_U c_A} \\
= & \exp\left\{-\frac{1}{2}\text{tr}F'F\right\} p(\Lambda, F) 2^k \frac{(d\Lambda)(dF)}{c_N c_U c_A} \\
\text{where } c_A = & \int |A|^{(T-n-k-1)/2} \exp\left\{-\frac{1}{2}\text{tr}A\right\} (dA) = 2^{(T-n)k/2} \Gamma_k\left(\frac{T-n}{2}\right).
\end{aligned}$$

We can summarize the transformations used to this point as

$$\begin{aligned}
\Pi &= U_1 S_1 V_1' \\
&= U_1 C' C S_1 V_1' \\
&= U \Lambda^* \\
&= U \kappa \kappa^{-1} \Lambda^* \\
&= F \Lambda
\end{aligned}$$

The resulting joint prior distribution for F and Λ is given by

$$\begin{aligned}
p(\Lambda, F)(d\Lambda)(dF) &= \exp\left\{-\frac{1}{2}\text{tr}F'F\right\} \exp\left\{-\frac{c_\lambda}{2}\text{tr}\Lambda'F'F\Lambda\right\} \frac{(d\Lambda)(dF)}{c} \\
c &= 2^{-k} c_N c_U c_A.
\end{aligned}$$

The resulting conditional priors have convenient normal forms such that they will be conjugate with the usual specification for the model for y . That is, the conditional prior for $\lambda = \text{vec}(\Lambda) | F$ is normal with zero mean and covariance matrix $\underline{V}_\lambda = I_n \frac{1}{c_\lambda} \otimes (F'F)^{-1}$. The conditional prior for $f = \text{vec}(F) | \Lambda$ is normal with zero mean and covariance matrix $\underline{V}_F = [I_k + c_\lambda \Lambda \Lambda']^{-1} \otimes I_T$. For the chosen transformations and distributions for the unidentified parameters, C and κ , we have results from Liu and Wu (1999) ensuring the sampler will converge. Specifically, the transformations we use form locally compact groups and the priors for the expanding parameters correspond to Haar measures. Further, the expanding parameters are independent of the identified parameters.

Parameter expansions have been used in earlier work in factor models to produce more efficient and simple sampling schemes (see, for a recent example, Ghosh and Dunson, 2009) and to accelerate the EM algorithm in factor models (Liu, Rubin and Wu, 1998). The approach in this paper is an application of that developed in Liu (1994) and Liu and Wu (1999) and shares some of the

features of Ghosh and Dunson (2009). However, a contribution of this paper that distinguishes it from this earlier work is to use parameter expansion to also achieve invariant inference. This builds upon earlier work on estimation of reduced rank models (Koop, Léon-González and Strachan (2010) & (2013)) which is natural as the factor model can be represented as a particular type of reduced rank regression model. However, relative to the standard reduced rank regression model, the factor model has some unique features that complicate estimation.

3.1 The full conditional posteriors

In this section we extend the model to allow exogenous variables and provide priors for the other parameters in the model to obtain the conditional posterior distributions that are used in the sampling algorithm.

The static factor model is often specified with exogenous variables, X . After the parameter expansions in the previous section we obtained the matrix of factors F and the loading matrix Λ . The model can now be written as

$$\begin{aligned} Y &= X\beta + F\Lambda + \varepsilon \text{ where } \varepsilon \sim N(0, H^{-1} \otimes I_T) \\ f &= \text{vec}(F) | \lambda \sim N(0, \underline{V}_F) \text{ and } \lambda = \text{vec}(\Lambda) | f \sim N(0, \underline{V}_\lambda) \\ b &= \text{vec}(\beta) \sim N\left(0, \Sigma_\beta \otimes (X'X)^{-1} \frac{1}{c_\beta}\right) \end{aligned}$$

where $\underline{V}_F = [I_k + c_\lambda \Lambda \Lambda']^{-1} \otimes I_T$ and $\underline{V}_\lambda = I_n \frac{1}{c_\lambda} \otimes (F'F)^{-1}$. We assume each diagonal element of $H = \text{diag}\{h_i\}$ has a gamma prior

$$p(h_i) \propto h_i^{\frac{\nu-2}{2}} \exp\left\{-\frac{h_i \underline{\nu}}{2\underline{\mu}}\right\} (dh_i)$$

such that

$$p(H) \propto |H|^{\frac{\nu-2}{2}} \exp\left\{-\frac{\underline{\nu}}{2\underline{\mu}} \text{tr} H\right\} (dH)$$

The likelihood function can be written as

$$L \propto |H|^{T/2} \exp\left[-\frac{1}{2} \text{tr} H (Y - X\beta - F\Lambda)' (Y - X\beta - F\Lambda)\right]$$

so that the conditional posteriors can be readily derived. First define

$$a = \begin{bmatrix} \text{vec}(\beta) \\ \text{vec}(\Lambda) \end{bmatrix} = \begin{bmatrix} b \\ \lambda \end{bmatrix}, \quad W = [(I_T \otimes X) \quad (I_k \otimes F)]$$

and $f = \text{vec}(F)$. Vectorizing the $T \times n$ error matrix ε gives the useful linear forms for f and a . Let $y = \text{vec}(Y)$, then let $x = (I_n \otimes X)$ and $\omega = (\Lambda' \otimes I_T)$,

$$\begin{aligned} \text{vec}(Y - X\beta - F\Lambda) &= y - (I_n \otimes X)b - (\Lambda' \otimes I_T)f \\ &= \tilde{y} - \omega f \text{ where } \tilde{y} = y - xb \\ &= y - (I_n \otimes X)b - (I_k \otimes F)\lambda \\ &= y - Wa. \end{aligned}$$

As the vectors f and a have normal priors and enter the likelihood linearly (conditional on the other parameters) the conditional posteriors result from standard computations. Specifically, the conditional posteriors have the following forms:

$$\begin{aligned} f|\beta, \Lambda, H, Y &\sim N(\bar{f}, \bar{V}_F) \\ a|F, H, Y &\sim N(\bar{a}, \bar{V}_a) \\ h_i|a, F &\sim G(\bar{\mu}_i, \bar{\nu}) \end{aligned}$$

where

$$\begin{aligned} \bar{f} &= \bar{V}_F (\Lambda H \otimes I_T) (y - xb) \\ \bar{V}_F &= [\Lambda (H + c_\lambda I_k) \Lambda' + I_k]^{-1} \otimes I_T, \\ \bar{a} &= \bar{V}_a W' (H \otimes I_T) y, \\ \bar{V}_a &= [W' (H \otimes I_T) W + \underline{V}_a^{-1}]^{-1} \\ \underline{V}_a^{-1} &= \begin{bmatrix} \Sigma_\beta^{-1} \otimes X' X c_\beta & 0 \\ 0 & c_\lambda I_k \otimes F' F \end{bmatrix} \\ y &= \text{vec}(Y), \\ W &= [I_n \otimes X \quad I_n \otimes F], \\ \bar{\nu} &= T + \underline{\nu}, \\ \bar{\mu}_i &= \varepsilon_i' \varepsilon_i + \frac{\underline{\nu}}{\underline{\mu}} \text{ and } \varepsilon = (\varepsilon_1, \dots, \varepsilon_n). \end{aligned}$$

An important question in factor models is the number of factors, k . With the parameter expansion specification above, we are able to use the Savage-Dickey density ratio (SDDR) to compute the Bayes factors for $k = 0$ to $k = k^*$, B_{0,k^*} (Verdinelli and Wasserman (1995)). The model with no factors occurs at the point $\Lambda = 0$ and at this point the factor are excluded from the likelihood. Under the priors we propose, the SDDR is computed as the ratio of the marginal posterior for Λ at the point $\Lambda = 0$ and the marginal prior for Λ evaluated at the same point. That is

$$SDDR = B_{0,k^*} = \frac{m_0}{m_{k^*}} = \frac{p(\Lambda = 0|y)}{p(\Lambda = 0)}$$

where m_{k^*} is the marginal likelihood for the model with k^* factors.

With a sequence of S draws from the posterior, we can compute the conditional posterior $p(\Lambda|\beta, F, H, y)$ and the conditional prior $p(\Lambda|F)$ at $\Lambda = 0$ to estimate the required ratio as:

$$\hat{B}_{0,k^*} = \frac{\frac{1}{S} \sum_{i=1}^S p(\Lambda = 0|\beta^{(i)}, F^{(i)}, H^{(i)}, y)}{\frac{1}{S} \sum_{i=1}^S p(\Lambda = 0|F^{(i)})},$$

where $i = 1, \dots, S$ indicates the draws from the sampler after the burn-in. In our results we report the log of the Bayes factor. Alternatively, we can report

posterior probabilities. From \widehat{B}_{0,k^*} we can compute the posterior probabilities for each model from the prior probabilities $p(k = k^*)$ and the relation

$$\frac{p(k = 0|y)}{p(k = k^*|y)} = B_{0,k^*} \frac{p(k = 0)}{p(k = k^*)}.$$

4 Empirical application

In this section present an empirical example to demonstrate the effect of lack of invariance and that our approach achieves invariance. We also present evidence on the efficiency of the sampling algorithm.

We use data on 6 daily international currency exchange rates relative to U.S. dollar over a time period of 1045 business days beginning in January 2007 and ending in December 2010. The returns are computed as $y_{it} = 100(p_{it}/p_{i,t-1} - 1)$, where p_{it} denotes the daily closing spot rate. The six series are the Australian Dollar (AUD), Euro (EURO), South Korean Won (KRW), Japanese Yen (JPY), Canadian Dollar (CAD), and British Pound (GBP). These represent some of the most heavily traded currencies over the period. The returns are plotted in Figure 1.

We fit the data using the following static factor model:

$$y_t = \mu + f_t \Lambda + \varepsilon_t \quad \varepsilon_t \sim iidN(0, \Sigma)$$

where y_t is the vector of six observed currency returns, $f_t \sim iidN(0, I_k)$ is a $1 \times k$ vector of unobserved factors, Λ is an $k \times n$ matrix of factor loadings. We first impose the identification assumption that Λ is upper triangular with positive diagonal elements:

$$\Lambda = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \vdots & a_k & \vdots & a_n \\ 0 & a_{22} & a_{32} & \vdots & a_{2k} & \vdots & a_{2n} \\ 0 & 0 & a_{33} & \vdots & \cdots & \vdots & a_{3n} \\ \cdots & \cdots & \cdots & \ddots & a_{k,k-1} & \vdots & \cdots \\ 0 & 0 & 0 & \vdots & a_{k^2} & \vdots & a_{kn} \end{pmatrix},$$

where $a_{ii} > 0$ for $i = 1, \dots, k$. We then compute the marginal likelihoods for four models. Using the ordering (AUD, EUR, KRW, JPY, CAD, GBP), we compute a single-factor model and then a two-factor model. Next, with the ordering (AUD, KRW, EUR, JPY, CAD, GBP), we again compute a single-factor model and then a two-factor model.

The log marginal likelihoods are computed via the Chib's method using 100 parallel chains each of length 50000. The results are reported in Table 3. As

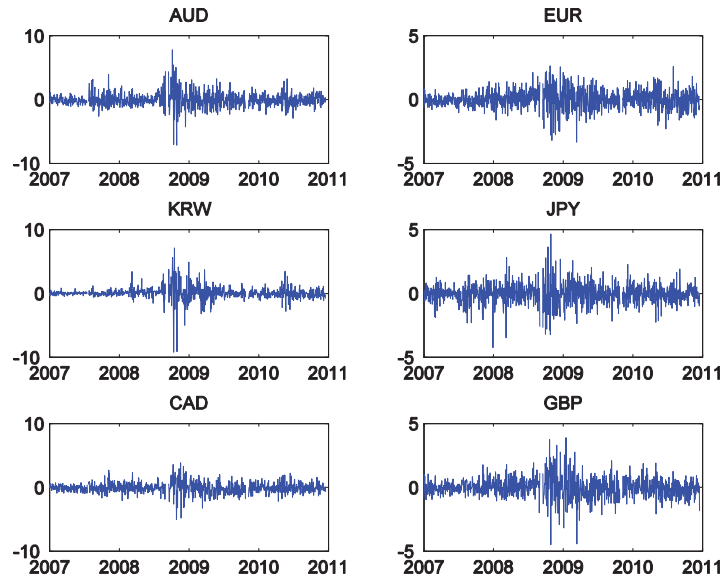


Figure 1: Daily returns of the six currencies.

the two marginal likelihoods for the models with one factor are almost the same (i.e., ordering made no difference) we only report one of these. The computed marginal likelihoods for the two-factors provide striking evidence of the effect of reordering. The log marginal likelihoods differ by about 142.

We can compute the log Bayes factors as the difference between the log marginal likelihoods. Under the ordering (AUD, KRW, EUR, JPY, CAD, GBP), there is a very strong preference for the two factor model over the one factor model with a log Bayes factor of -63.6. However, under the ordering (AUD, EUR, KRW, JPY, CAD, GBP), there is a very strong preference for the one factor model with the log Bayes factor of 78.3. The reordering of the variables has shifted the evidence on the number of factors in the opposite direction. The invariant specification selects a model with two factors over the one factor model with a log Bayes factor of -309.6.

Table 3: Log marginal likelihoods and the corresponding numerical standard errors for the competing models.

Model	log marginal likelihood	numerical standard error
1-factor (AUD, KRW, EUR, JPY, CAD, GBP)	-7572.9	3.40
2-factor (AUD, EUR, KRW, JPY, CAD, GBP)	-7636.5	5.64
2-factor (AUD, KRW, EUR, JPY, CAD, GBP)	-7494.6	1.16

Table 3 shows the numerical standard errors (NSEs) for the estimated log marginal likelihoods under each ordering. The NSEs for the one and two factor models for the invariant model are 0.23 and 0.25 respectively. These results suggest, anecdotally at least, that the estimation of the log marginal likelihoods is much more efficient under the invariant specification. The NSE for the one factor model reduces from 3.4 to 0.23, a reduction of over 93%. The NSE for the two factor model reduces from 5.64 for the first ordering, and 1.16 for the second ordering to 0.25, a reduction of over 95% and 78%.

Figure 2 reports the inefficiency factors for blocks of parameters from the model estimated with the Geweke and Zhou specification (GZ) and the parameter expanded invariant model (PX). The inefficiency measures give an estimate of the number of draws needed to have as much information about the posterior as we would obtain from one independent draw. The smaller the inefficiency factor the better is the sampler. It is clear that the parameters are generally much more efficiently estimated using the expanded parameter specification. These results are consistent with those found in Ghosh and Dunson (2009). The distributions of both the loading matrix, Λ , and the factors, F , are less dispersed and lower for the parameter expanded model, but this is also true for the idiosyncratic variances, H , and the exogenous variables coefficients, β .

5 Concluding remarks

In this paper, we propose a specification for the static factor model that requires no ordering restrictions and so the choice of number of factors cannot depend upon the chosen ordering. By augmenting the posterior with a number of unidentified parameters with appropriate priors, the model can be computed using standard distributions and the draws are relatively efficient.

The specification we propose nests many of the existing and popular specifications used in factor analysis. Thus each of these specifications are attainable directly from the output from estimating our specification.

Although for convenience we have only considered the static factor model, this approach is readily extended to allow dynamics in the state equation. Such an extension would involve using an informative prior on the space $sp(F)$ such as the orthogonal projective Gaussian distribution as used in Koop, León-González and Strachan (2011). This would involve transforming from F to $F_c = RF$

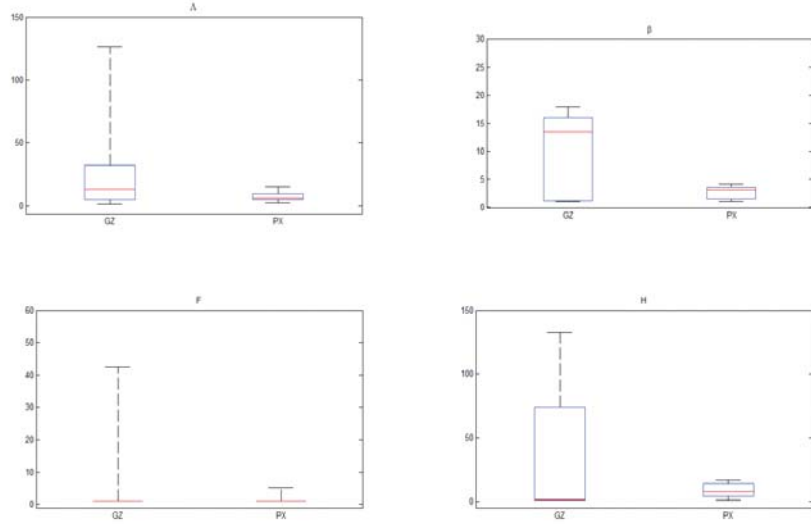


Figure 2: Box and whisker plots of inefficiency factors for blocks of parameters: (Λ, β, F, H) . The models were estimated using Gibbs sampling. *GZ* refers to the model with the upper triangular Λ_1 with positive elements on the main diagonal, and *PX* refers to the parameter expansion of the invariant specification.

where the matrix R captures the dynamics in F_c . For example, R may be a differencing matrix such that the rows of F follow a random walk as

$$f_{c,t} = f_{c,t-1} + f_t = \sum_{i=1}^t f_i \text{ and so}$$

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Alternatively $R = R(\rho)$ may be a function of parameters to allow richer dynamics such as in autoregressive processes. For example, an $AR(1)$ state equation

$$f_{c,t} = \rho f_{c,t-1} + f_t = \sum_{i=1}^t \rho^{t-i} f_i \text{ and so}$$

$$R(\rho) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \rho & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ \rho^{t-1} & \rho^{t-2} & \cdots & 1 \end{bmatrix}.$$

The full matrix of factors then becomes $F_c = R(\rho) F$ which implies a prior for a process with a zero mean process with covariance matrix

$$(I_k + c_\lambda \Lambda \Lambda')^{-1} \otimes (R(\rho)' R(\rho))^{-1}.$$

Another implication of the invariance specification, which we have not exploited in this paper, is that we are able to compute the Bayes factors for the number of factors using the Savage-Dickey density ratio. This approach requires only the conditional posterior and the conditional prior for Λ . This greatly simplifies the computation of the posterior probabilities. This approach cannot be used in the non-invariant specifications as they exclude the point $\Lambda = 0$ from the support of the loading matrix parameter.

In computing the models in this paper, it became evident that there is a relationship between the computational efficiency and accuracy of marginal likelihood estimates, and the proximity of the posterior to the point of discontinuity. Models that are specified such that the posterior is invariant to reordering tend to have lower numerical standard errors. The accuracy of estimation of the marginal likelihood plays an important role in the confidence we have in the conclusions we make. This relationship is a topic of current research.

6 References

- Aßmann, Boysen-Hogrefe, and Pape (2012) “The Directional Identification Problem in Bayesian Factor Analysis: An Ex-Post Approach” Working paper 2012-11, Department of Economics, Christian-Albrechts-Universität Kiel, Germany
- Aguilar, O. and M. West (2000) “Bayesian Dynamic Factor Models and Portfolio Allocation,” *Journal of Business & Economic Statistics*, Vol. 18, No. 3, 338-357.
- Bai, J. and Ng, S. (2002) “Determining the number of factors in approximate factor models,” *Econometrica*, Vol. 70, No. 1, 191-221.
- Chib, S., Nardari, F., and N. Shephard (2006) “Analysis of high dimensional multivariate stochastic volatility models,” *Journal of Econometrics* 134 (2006) 341–371.
- Connor, G. and R. Korajczyk (1986) “Performance measurement with the arbitrage pricing theory: A new framework for analysis,” *Journal of Financial Economics* 15, 373-394.
- Forni M., Giannone D., Lippi M. and L. Reichlin (2009) “Opening the Black Box: Structural Factor Models with large cross-sections”, *Econometric Theory* 25, 1319-1347.
- Frühwirth-Schnatter, S. and H. F. Lopes (2010) “Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown” Technical Report, Chicago Booth <http://faculty.chicagobooth.edu/hedibert.lopes/research/reports.html>

- Geweke, J. (1996) "Bayesian reduced rank regression in econometrics," *Journal of Econometrics* 75, 121-146.
- Geweke, J., and Zhou, G. (1996) "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557-587.
- Ghosh, J., and Dunson, D.B. (2009) "Default priors and efficient posterior computation in Bayesian factor analysis," *Journal of Computational and Graphical Statistics* 18, 306-320.
- James, A. T. (1954) "Normal multivariate analysis and the orthogonal group," *Annals of Mathematical Statistics*, 25, 40-75.
- Kaufmann and Schumacher (2012) "Bayesian estimation of sparse dynamic factor models with order-independent identification" Working paper 13.04, Study Center Gerzensee, Foundation of the Swiss National Bank.
- Koop G., R. Léon- González and R. W. Strachan (2010) "Efficient posterior simulation for cointegrated models with priors on the cointegration space" *Econometric Reviews* Volume 29, Issue 2, 224-242.
- Koop G., R. Léon- González and R. W. Strachan (2011) "Bayesian Inference in the Time Varying Cointegration Model," *The Journal of Econometrics* 165, 210-220.
- Koop G., R. Léon-Gonzalez and R. W. Strachan (2013) "Bayesian Model Averaging in the Instrumental Variable Regression Model" forthcoming in *The Journal of Econometrics*.
- Liu, J.S. (1994), "The collapsed Gibbs sampler with applications to a gene regulation problem," *Journal of the American Statistical Association* 89, 958-966.
- Liu C., D. B. Rubin, Y. N. Wu (1998) "Parameter Expansion to Accelerate EM: The PX-EM Algorithm," *Biometrika*, Vol. 85, No. 4, pp. 755-770
- Liu, J. and Wu, Y. (1999) "Parameter expansion for data augmentation," *Journal of the American Statistical Association*, 94, 1264-1274.
- Lopes H. F. and M. West (2004) "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica* 14, 41-67.
- Magnus, J.R., Neudecker, H. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.
- Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. New York: Wiley.
- Schott, J.R. (1997) *Matrix Analysis for Statistics*. Wiley, New York.
- Strachan, R. and Inder, B. (2004) "Bayesian analysis of the error correction model," *Journal of Econometrics* 123, 307-325.

Verdinelli, I. and L. Wasserman (1995) “Computing Bayes Factors using a generalization of the Savage-Dickey Density Ratio,” *Journal of the American Statistical Association* 90, 614-618.

Villani, M. (2005) “Bayesian reference analysis of cointegration,” *Econometric Theory* 21, 326-357.

7 Appendix I

Proof of Theorem (1).

Denote by Λ^i the $k \times n$ loading matrix under order i . Partition Λ^i , after imposing one of the Bayesian invariant identifying restrictions, as follows:

$$\Lambda^i = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ 0 & \Lambda_{22} & \Lambda_{23} \end{bmatrix}$$

where Λ_{11} is a $k_1 \times k_1$ matrix, Λ_{22} and Λ_{23} are both $k_2 \times k_2$ where $k_1 + k_2 = k$. Under \mathfrak{T}^+ , $\Lambda_{11} = \Lambda_{11}^+$ and $\Lambda_{22} = \Lambda_{22}^+$ are lower triangular matrices with positive diagonal elements. Under \mathfrak{T}^1 , $\Lambda_{11} = \Lambda_{11}^1$ and $\Lambda_{22} = \Lambda_{22}^1$ are lower triangular matrices with ones on the diagonal.

The ordering is imposed by the structure of the first k columns. We consider changing the order of the variables such that the ordering is imposed by the structure of the first k_1 columns and the last k_2 columns. Thus we will transform Λ_{23} to meet the identifying restrictions. This will involve transforming from (F, Λ) to $(\tilde{F}, \tilde{\Lambda})$ with the appropriate definition of the $k_2 \times k_2$ matrix G in

$$\begin{aligned} F\Lambda &= F \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ 0 & \Lambda_{22} & \Lambda_{23} \end{bmatrix} \\ &= F \begin{bmatrix} I_{k_1} & 0 \\ 0 & G^{-1} \end{bmatrix} \begin{bmatrix} I_{k_1} & 0 \\ 0 & G \end{bmatrix} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ 0 & \Lambda_{22} & \Lambda_{23} \end{bmatrix} \\ &= \tilde{F}\tilde{\Lambda} \text{ where} \\ \tilde{F} &= F \begin{bmatrix} I_{k_1} & 0 \\ 0 & G^{-1} \end{bmatrix} \text{ and} \\ \tilde{\Lambda} &= \begin{bmatrix} I_{k_1} & 0 \\ 0 & G \end{bmatrix} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ 0 & \Lambda_{22} & \Lambda_{23} \end{bmatrix} \\ &= \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ 0 & \Lambda_{22}^j & \Lambda_{23}^j \end{bmatrix}. \end{aligned}$$

Finally, to move the columns of $\tilde{\Lambda}$ such that the identifying restrictions are

placed upon the first k columns of the loading matrix, we reorder by

$$\begin{aligned} \tilde{\Lambda} \begin{bmatrix} I_{k_1} & 0 & 0 \\ 0 & 0 & I_{k_2} \\ 0 & I_{k_2} & 0 \end{bmatrix} &= \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ 0 & \Lambda_{22}^j & \Lambda_{23}^j \end{bmatrix} \begin{bmatrix} I_{k_1} & 0 & 0 \\ 0 & 0 & I_{k_2} \\ 0 & I_{k_2} & 0 \end{bmatrix} \\ &= \begin{bmatrix} \Lambda_{11} & \Lambda_{13} & \Lambda_{12} \\ 0 & \Lambda_{23}^j & \Lambda_{22}^j \end{bmatrix} = \Lambda^j. \end{aligned}$$

For reordering within the space \mathfrak{T}^+ , we first let $\Lambda_{23} = C\Lambda_{23}^+$ be the QR decomposition of Λ_{23} where $C'C = I_{k_2}$. The change in order implies $G = C'$ and $G^{-1} = C$. For the space \mathfrak{T}^1 , the change in order implies $G = D^{-1}C'$ and $G^{-1} = CD$ where $D = \text{diag}\{\lambda_{23,11}, \lambda_{23,22}, \dots, \lambda_{23,k_2k_2}\}$ is a diagonal matrix and $\lambda_{23,ii}$ is the i^{th} diagonal element of Λ_{23}^+ .

From these transformations we can demonstrate the implications of reordering and how $|\Lambda_1| = 0$ implies a discontinuity in this transformation. Consider first the case of transforming from the space $\mathfrak{T}^{+,i}$ to the space $\mathfrak{T}^{+,j}$. If we use the normalization with positive diagonal elements then after reordering we wish to map back into the space \mathfrak{T}^+ . Now consider the points where $|\Lambda_{23}| = 0$. In this case the QR decomposition, $\Lambda_{23} = C\Lambda_{23}^+$, will result in $|\Lambda_{23}^j| = |\Lambda_{23}^+| = 0$ as Λ_{23}^+ will have some zeros on the diagonal as it has reduced rank. Therefore, at the point $|\Lambda_{23}^j| = |\Lambda_{23}^+| = 0$ the matrix Λ^j cannot meet the identifying restrictions, that is $\Lambda^j \notin \mathfrak{T}^{+,j}$. The transformation is discontinuous and maps outside the original space such that we cannot claim invariance.

The above result implies that within the space \mathfrak{T}^1 , the change in order implies D will have zeros on the diagonal such that G is not defined. This collection of transformations, therefore, does not form a matrix or linear group and, again, we cannot claim invariance.

We can demonstrate an alternative argument for lack of invariance by considering that the transformations implied by the transformations do not map from a space onto itself. We show this in terms of the space spanned by the loading matrix.

Each of the restrictions under $\mathfrak{T}^{m,i}$ $m \in \{+, 1\}$ implies that each of the k ($n \times 1$) vectors in Λ' are tied to k coordinate axes such that $\text{sp}(\Lambda')$ cannot lie in the space spanned by the remaining $(n - k)$ coordinate axes. This is achieved by forcing k rows of Λ' , Λ'_1 , to be full rank. For example, the vectors may be tied to the first k axes. The k -dimensional space of the matrix made up of the first k coordinate axes is

$$\mathfrak{p}_I = \text{sp} \begin{pmatrix} I_k \\ 0 \end{pmatrix} \in G_{k,n-k}$$

where $G_{k,n-k}$ is the Grassmann manifold, and the $(n - k)$ -dimensional space orthogonal to \mathfrak{p}_I is $\mathfrak{p}_{I\perp} \in G_{n-k,k}$. If $k \geq n - k$, then let \mathfrak{p}_\perp be any point in $G_{k,n-k}$ such that $\mathfrak{p}_{I\perp} \subseteq \mathfrak{p}_\perp$, and if $n - k \geq k$, then let \mathfrak{p}_\perp be any point in $G_{n-k,k}$ such that $\mathfrak{p}_\perp \subseteq \mathfrak{p}_{I\perp}$. The support of the space $\mathfrak{p}_i = \text{sp}(\Lambda')$ for all $\Lambda \in \mathfrak{T}^{m,i}$ is a

subset of the Grassmann manifold, $\mathbf{p}_i \in S_i \subset G_{k,n-k}$, since $\mathbf{p}_\perp \notin S_i$ is excluded. That is, $G_{k,n-k} \equiv S_i \cup \mathbf{p}_\perp$ and S_i and \mathbf{p}_\perp are disjoint. However, $\mathbf{p}_I \in S_i$.

Any other ordering of the variables $\mathfrak{T}^{m,j}$ will have support S_j such that $\mathbf{p}_\perp \in S_j$. Invariance requires that there exists a group of transformations from the space, in this case S_i , into itself. This group will be one to one of S_i onto S_i . However, reordering involves changing the support of \mathbf{p}_i from S_i to S_j .

8 Appendix II

In this we derive the Savage-Dickey density ratio (SDDR) for a model with $k = k^*$ factors to a model with $k = k^* - 1$ factors by computing the marginal posterior and prior for the smallest singular value, s_k , of the loading matrix. If there is $k - 1$ rather than k factors in the model, then the $(k \times n)$ matrix Λ will have rank $k - 1$. This restriction will occur at the point $s_k = 0$.

To obtain these distributions, we begin with the conditional densities for Λ

$$p(\Lambda|\theta, y) \text{ and } p(\Lambda|\theta)$$

where θ is a vector of all other parameters in the model. Next, transform from the loading matrix, Λ , to parameterize the restriction to lower rank as an invariant function of the smallest singular value, s_k . The transformation will take the form

$$\Lambda = \alpha\delta' + \alpha_\perp\sigma\delta'_\perp$$

in which α_\perp and δ_\perp are unit vector analytic functions of $sp(\alpha)$ and $sp(\delta)$ respectively and $\alpha'\alpha = I_{k-1}$. The matrix dimensions are α ($k \times k - 1$), δ ($n \times k - 1$), α_\perp ($k \times 1$), σ (1×1), and δ_\perp is ($n \times 1$). At the point $s_k = 0$ then $\sigma = 0$ and so the SDDR is computed as

$$\begin{aligned} SDDR &= \frac{\int p(\alpha, \delta, \sigma = 0|y) d(\alpha, \delta)}{\int p(\alpha, \delta, \sigma = 0) d(\alpha, \delta)} \\ &= \frac{p(\sigma = 0|y)}{p(\sigma = 0)}. \end{aligned}$$

To obtain the values $p(\sigma = 0|y)$ and $p(\sigma = 0)$, we approximate the integrals by Rao-Blackwellizing. That is, with M draws from the posterior and N draws from the priors of α and δ , we compute

$$SDDR = \frac{\sum_{i=1}^M p(\sigma = 0|\alpha^{(i)}, \delta^{(i)}, \theta^{(i)}, y)}{\sum_{j=1}^N p(\sigma = 0|\alpha^{(j)}, \delta^{(j)}, \theta^{(j)})}.$$

Thus we need expressions for the conditional posterior $p(\sigma = 0|\alpha, \delta, \theta, y)$ and the conditional prior $p(\sigma = 0|\alpha, \delta, \theta)$.

Here is present how the transformation from Λ to (α, δ, σ) is achieved. Let the singular value decomposition of Λ be represented by

$$\begin{aligned}
\Lambda &= USV' \\
&= [U_1 \quad u_k] \begin{bmatrix} S_1 & 0 & 0 \\ 0 & s_k & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ v_k' \\ V_3' \end{bmatrix} \\
&= U_1 S_1 V_1' + u_k s_k v_k' \\
U &= [U_1 \quad u_k] \in O(k), \quad V = [V_1 \quad v_k \quad V_3] \in O(n) \\
S &= \begin{bmatrix} S_1 & 0 & 0 \\ 0 & s_k & 0 \end{bmatrix}
\end{aligned}$$

where S_1 is a diagonal matrix containing the $k-1$ largest singular values of Λ and the scalar s_k is the smallest singular value. Given Λ is an $k \times n$ matrix, this defines the dimensions of all matrices and submatrices above. Next we introduce an unidentified $(k-1) \times (k-1)$ orthogonal matrix $T \in O(k-1)$ with uniform distribution on $O(k-1)$ via

$$\begin{aligned}
U_1 S_1 V_1' &= U_1 S_1 T T' V_1' = \alpha \delta' \\
\alpha &= U_1 S_1 T \text{ and } \delta = V_1 T.
\end{aligned}$$

Next, introduce two more unidentified parameters: $\tau_\alpha, \tau_\delta \in O(1) = \{-1, 1\}$. Then transform the product $u_k s_k v_k' = u_k \tau_\alpha \tau_\delta s_k \tau_\delta \tau_\delta v_k' = \alpha_\perp \sigma \delta_\perp'$ where $\alpha_\perp = u_k \tau_\alpha$, $\sigma = \tau_\alpha s_k \tau_\delta$, and $\delta_\perp = v_k \tau_\delta$. Thus the restriction that Λ has rank $k-1$, so that there is one less factor in the model, is parameterized by $\sigma = 0$. Importantly for computation, the Jacobian $J(\Lambda, T, \tau_\alpha, \tau_\delta \rightarrow \alpha, \delta, \sigma)$ is not a function of σ and, at $\sigma = 0$, is not a function of α_\perp or δ_\perp .

To compute the SDDR, we need the conditional distribution for σ given (α, δ) . We have the conditional posterior for $\lambda = \text{vec}(\Lambda)$ (marginal of β) as

$$\lambda | F, H, Y \sim N(\bar{\lambda}, \bar{V}_\lambda)$$

where

$$\begin{aligned}
\bar{\lambda} &= \bar{V}_\lambda (H \otimes F') y, \\
\bar{V}_\lambda &= (H + c_\lambda I_k)^{-1} \otimes (F' F)^{-1}.
\end{aligned}$$

Vectorizing Λ we have with $d = \text{vec}(\delta)$,

$$\begin{aligned}
\lambda &= \text{vec}(\Lambda) = (I_n \otimes \alpha) d + (\delta_\perp \otimes \alpha_\perp) \sigma \text{ and so} \\
\sigma &= (\delta_\perp' \otimes \alpha_\perp') [\lambda - (I_n \otimes \alpha) d] \\
&= [(\delta_\perp' \otimes \alpha_\perp') \lambda - (\delta_\perp' \otimes \alpha_\perp' \alpha) d] \\
&= (\delta_\perp' \otimes \alpha_\perp') \lambda.
\end{aligned}$$

The conditional posterior distribution of σ is then

$$\sigma | \alpha, \delta, F, H, Y \sim N(\bar{\sigma}, \bar{V}_\sigma)$$

where

$$\begin{aligned}\bar{\sigma} &= (\delta'_{\perp} \otimes \alpha'_{\perp}) \bar{\lambda}, \\ \bar{V}_{\sigma} &= (\delta'_{\perp} \otimes \alpha'_{\perp}) \bar{V}_{\lambda} (\delta_{\perp} \otimes \alpha_{\perp}).\end{aligned}$$

In a similar way, the conditional prior for σ is derived from the prior for Λ and draws are obtained from the a Gibbs sampler consisting of the conditional densities for f and λ at the beginning of Section 4.

The Conditional Prior

The prior distribution for $\lambda = \text{vec}(\Lambda)$ is

$$\lambda = \text{vec}(\Lambda) | f \sim N(0, \underline{V}_{\lambda})$$

where $\underline{V}_{\lambda} = I_n \frac{1}{c_{\lambda}} \otimes (F'F)^{-1}$.

The conditional prior distribution of σ is then

$$\sigma | \alpha, \delta, F \sim N(0, \underline{V}_{\sigma})$$

where $\underline{V}_{\sigma} = (\delta'_{\perp} \otimes \alpha'_{\perp}) \underline{V}_{\lambda} (\delta_{\perp} \otimes \alpha_{\perp})$. The Gibbs sampler for drawing λ uses the conditionals

$$f = \text{vec}(F) | \lambda \sim N(0, \underline{V}_F) \quad \text{and} \quad \lambda = \text{vec}(\Lambda) | f \sim N(0, \underline{V}_{\lambda})$$

where $\underline{V}_F = [I_k + c_{\lambda} \Lambda \Lambda']^{-1} \otimes I_T$ and $\underline{V}_{\lambda} = I_n \frac{1}{c_{\lambda}} \otimes (F'F)^{-1}$.