A Metropolis-Hastings algorithm for reduced rank covariance matrices with application to Bayesian factor models

GAETANO CARMECI

*D Dipartimento di Scienze Economiche e Statistiche, Università di Trieste, Piazzale Europa 1, I-34127 Trieste. E-mail: carmeci@mail.univ.trieste.it
Abstract

Most of the proposed Markov chain Monte Carlo (MCMC) algorithms for estimating static and dynamic Bayesian factor models are parametrized in terms of the loading matrix and the latent common factors which are sampled into two separate blocks. In this paper, we propose a novel implementation of the MCMC algorithm which is designed for the model parametrized in terms of the reduced rank covariance matrix underlying the factor model. Hence, the strategy proposed makes it possible to sample directly from the reduced rank covariance matrix. The alternative parameterization of the model is undoubtedly more natural for the linear dynamic factor model. Furthermore, it allows us to rewrite the static factor model as a hierarchical (multilevel) linear model. In this way, a better mixing of the Markov chain is obtained. We adopt, as prior for the singular covariance matrix, the noninformative prior distribution first considered by Diaz-Garcia and Gutierrez (2006). We implement an efficient MCMC algorithm characterized by the sampling of the singular covariance matrix and the associated (unobserved) systematic component in one block. Furthermore, we propose the sampling of the singular covariance matrix marginalized over the systematic component. For this purpose, we develop a Metropolis-Hastings (M-H) step that takes explicitly into account the curved geometry of the support of the target distribution. The proposal distribution is based on a mixture of Wishart Singular distributions; see Diaz-Garcia et al. (1997). It is worth noting that, as a result of working with singular distributions, the prior and posterior densities, as well as the density of the proposal distribution in the M-H step, are specified with respect to Hausdorff measure and integral. The curved geometry of the support has implications for the Bayesian inference on the reduced rank covariance matrix too. That is, a Bayesian point estimator which preserves the original structure of the singular covariance matrix is required. We propose a Bayesian point estimator which is obtained by the generalized Choleski decomposition for reduced rank covariance matrices. We apply our approach to static factor models and present an empirical illustration on exchange rates. Moreover, we consider a simple but important example of linear dynamic factor model in time series analysis: the multivariate local level model with common trends; see Harvey and Koopman (1997). A Bayesian analysis of three monthly US short-term interest rates is presented.

KEYWORDS: Metropolis-Hastings algorithm, positive semi-definite covariance matrix, singular matrix, Wishart Singular distribution, Bayesian inference, Generalized Inverse Wishart distribution, Hausdorff measure, factor model, unobserved components, state space model, local level model, common trends, cointegration.

JEL classification: C15, C11, C32
1. Introduction

Factor models have a long tradition in social science; they are important in a wide range of subjects, including economics, finance, statistics and econometrics. Furthermore in latest years, as testified by the 2004’s special issue of the Annals in the *Journal of Econometrics* edited by Croux et. al. (2004), there has been renewed interest in dynamic factor models and related state space models, particularly in financial and macroeconometrics.

From the Bayesian viewpoint, Markov chain Monte Carlo (MCMC) methods for estimating static and dynamic factor models have been proposed, among others, by Geweke and Zhou (1996), Pitt and Shephard (1999b), Aguilar and West (2000), Jacquier et. al. (1995) and Lopes and West (2004). All these MCMC algorithms are parametrized in terms of the loading matrix and the latent common factors which are sampled into two separate blocks.

In this paper, we propose a novel implementation of the MCMC algorithm which is designed for the model parametrized in terms of the reduced rank covariance matrix underlying the factor model. Hence, the strategy proposed makes it possible to sample directly from the reduced rank covariance matrix.

Indeed, the reduced rank covariance matrix is at the heart of the factor model. For example, in classical factor analysis the covariance matrix of the linear combinations of common factors is a singular, positive semidefinite matrix; see, e.g., Anderson (2003). As an other simple but important example in time series analysis, let’s consider the multivariate random walk plus noise process, usually named as *local level* model; see, e.g. Harvey (1989) and Harvey and Shephard (1993). A (nonstationary) dynamic common factor model arises when the covariance matrix of the disturbances driving the trend component is of positive but reduced rank; see, e.g. Harvey and Koopman (1997). In this model the rank deficiency of the covariance matrix implies the existence of *common trends* in the time series, and in turn this means that the time series are *cointegrated*.

In the paper, we apply the proposed novel implementation of the MCMC algorithm to static and linear dynamic factor models. However, our approach can be fruitfully employed in all models in which the Bayesian inference on a random positive
The alternative parameterization of the model in terms of the reduced rank covariance matrix is undoubtedly more natural for the linear dynamic factor model. Furthermore, it allows us to rewrite the static factor model as a hierarchical (multilevel) linear model. In this way we obtain two significant advantages over the standard parameterization. First, in contrast to the standard parameterization where the loading matrix and the latent common factors enter the model as a product, in the proposed parameterization the singular unobserved components and their reduced rank covariance matrix appear separately. As a result, a better mixing of the Markov chain, which improves the overall efficiency of the algorithm, is obtained. Second, contrary to the standard formulation where a priori constraints must be imposed on the loading matrix to achieve parameter identification\(^1\), in the proposed formulation the identification of the parameters associated to the systematic component of the model requires fulfilment of the reduced rank condition only.

There is also a potential drawback to our approach: working with the reduced rank covariance matrix, rather than the loading matrix, increases the computational burden. Furthermore, the larger the dimension of the factor model, the higher will be such increment with respect to the computational burden required by the standard parameterization. Therefore, the proposed MCMC algorithm appears to be better suited for small and medium size factor models.

In a Bayesian context, the issue of appropriately choosing a prior for the unknowns in the model must be addressed at the outset. In our approach this is a crucial step because a singular covariance matrix results to be involved. The analysis about singular random matrices has received renewed attention in the literature.\(^2\) In this paper, we adopt, as prior for the singular covariance matrix, the noninformative prior distribution first considered by Diaz-Garcia and Gutierrez (2006) in the context of

---

\(^1\) For example, Geweke and Zhou (1996) and subsequently many other authors have specified the factor loading matrix as a full rank, block lower triangular matrix with diagonal elements strictly positive.

Bayesian singular multivariate linear models. While in the case of Bayesian singular multivariate linear models the joint posterior distribution for the parameters can be analytically obtained, this is not possible in the case of factor models, hence we must rely upon simulation methods. To this end, we implement an efficient MCMC algorithm characterized by the sampling of the singular covariance matrix and the associated (unobserved) systematic component in one block. Furthermore, we propose the sampling of the singular covariance matrix marginalized over the systematic component. For this purpose, we develop a Metropolis-Hastings (M-H) step that takes explicitly into account the curved geometry of the support of the target distribution. Indeed, the crucial issue in the design of the M-H step is that the support of the target distribution is not a vector space but a manifold. As such, a linear combination of any two elements of this set may not be an element of the same set. Therefore, the curvature of the support prevents us from using standard candidate-generating densities such as the random walk M-H chain; see, e.g., Chib (2001). Accordingly, we develop a simple alternative proposal distribution that resolves the issue and that, like the random walk M-H chain for linear spaces, depends upon the current location of the chain. Our proposal distribution is based on a mixture of Wishart Singular distributions. In the paper we show that, by fine tuning the setting parameters of this proposal distribution, we succeed in efficiently moving the chain through the whole support of the target density.

It is worth noting that, as explained by Diaz-Garcia et al. (1997) and, very recently, by Diaz-Garcia and Gutierrez (2006) and Diaz-Garcia (2007), [A note about measures and Jacobians of singular random matrices, Journal of Multivariate Analysis, 98, pgg. 960-969], the singular random matrices considered here possess a density only with respect to Hausdorff measure. Therefore, the prior and posterior densities, as well as the density of the proposal distribution in the M-H step, are specified with respect to

---

3 In the same context, Diaz-Garcia and Ramos-Quiroga (2003) have proposed a proper generalized natural conjugate prior density.

4 Diaz-Garcia and Gutierrez (2006) show that the marginal posterior distribution for the reduced rank covariance matrix is a central, Singular Generalized Inverse Wishart. This distribution is a generalization of the Inverse Wishart for rank-$n$ positive semidefinite $mxm$ symmetric matrices with $n$ distinct positive eigenvalues.

5 Our MCMC blocking strategy is similar to the approach recently proposed by Chib et al. (2006) in a related context, which involves a classical factor model. They show that a more efficient MCMC algorithm is obtained by sampling the loading matrix and the common factors in one block.

6 For a definition of the Wishart Singular distribution see, e.g., Diaz-Garcia et al. (1997).
Hausdorff measure and integral, instead of the usual Lebesgue measure and integral. However, we will show that, from a practical point of view, working with Hausdorff measure and integral does not require any special attention in the MCMC implementation.

Further, we notice that the curved geometry of the support has implications for the Bayesian inference on the reduced rank covariance matrix too. In fact, although posterior means for the parameters are routinely calculated as averages of the simulated samples, the averages thus obtained do not preserve the matrix structure of the unknown singular covariance matrix when the geometry of the support is curve.

To avoid misleading results, a Bayesian point estimator which preserves the original matrix structure is required. The paper suggests a solution to this problem: we propose a Bayesian point estimator which is obtained by the generalized Choleski decomposition for reduced rank covariance matrices. Moreover, we show how it is possible to obtain a Bayesian point estimator of the null space of the singular covariance matrix, as well as of its column space. It is interesting to notice that in the case of the local level model with common trends such null space coincides to the cointegration space. For these purposes, we exploit the results presented by Villani (2006) in the context of the Bayesian Vector Error Correction Model.

More to the point, the strategy will be extended to investigate the case when the number of common factors is unknown; hence the rank of the singular covariance matrix must be selected too. The rank selection will be based on the Bayesian Deviance Information Criterion (DIC), proposed by Spiegelhalter et al. (2002) as a simple alternative to Bayes factors or posterior odds computations.

The plan of the paper is as follows. In section two we present our approach for drawing singular covariance matrices in a different context from factor models. That is, for ease of presentation, but without loss of generality, we start the analysis making inference on a normal singular population. In section three we apply the proposed novel implementation of the MCMC algorithm to the static factor model and present an application to exchange rates. In section four we consider a multivariate local level model with common trends and describe how our MCMC algorithm has to be adapted. Moreover, we present an application of our approach to three monthly US short-term interest rates as empirical illustration. Section five concludes.
2. A Metropolis-Hastings algorithm for singular covariance matrices

For ease of presentation, but without loss of generality, we initially introduce our approach for drawing singular covariance matrices in a different context from factor models, i.e. we start the analysis making inference on a normal \( m \)-dimensional singular population. This model, which can be thought as the simplest Bayesian singular multivariate linear model\(^7\), is defined as follows. For the \( m \times 1 \) vector of observations \( y_t \), it is assumed that

\[
y_t = \eta_t, \quad \eta_t \sim SNID(0; \Sigma_\eta), \quad t = 1, \ldots, T,
\]

where \( SNID(0; \Sigma_\eta) \) denotes that the vector is serially independent and distributed as a singular Normal with zero mean vector and \( m \times m \) positive semidefinite covariance matrix \( \Sigma_\eta \) of known reduced rank \( n \) with \( n < m < T \).

This model can be written in matrix form as

\[
Y = \eta,
\]

where \( Y = (y_1, \ldots, y_T)' \) and \( P(\eta/\Sigma_\eta) \equiv N^{T,n}_{\text{ton}}(0, \Sigma_\eta \otimes I_T) \). As in Diaz-Garcia and Gutierrez (2006), we denote by \( N^{T,n}_{\text{ton}}(0, \Sigma_\eta \otimes I_T) \) the corresponding \( T \times m \) matrix-variate singular Normal distribution with \( \text{rank}(\Sigma_\eta) = n \) and we propose the following noninformative prior density for \( \Sigma_\eta \):

\[
dP(\Sigma_\eta) \propto \prod_{i=1}^{n} \lambda^{-\left(2m-n+1\right)/2} (d\Sigma_\eta)
\]

\(^7\) As pointed out in the introduction, MCMC methods are not required for this model, because by now the joint posterior distribution for the parameters of the Bayesian singular multivariate linear model has been analytically obtained by Diaz-Garcia and Gutierrez (2006) and Diaz-Garcia and Ramos-Quiroga (2003), under the assumption of a noninformative prior and a proper generalized natural conjugate prior respectively. Nevertheless, this model allows us to neatly set up the key ingredients of the proposed M-H step.
where \( \lambda_i, i = 1, \ldots, n \) are the non-null eigenvalues of \( \Sigma_q \). Diaz-Garcia and Gutierrez (2006) point out that in eq. (3) the measure \( (d\Sigma_q) \) is not the Lebesgue measure, but the Hausdorff measure defined on \( S_m^+(n) \), the manifold of dimension \( mn-n(n-1)/2 \) of the real symmetric rank-\( n \) positive semi-definite \( mxm \) matrices with \( n \) distinct positive eigenvalues. Moreover, by using the nonsingular part of the spectral decomposition for \( \Sigma_q \), the Hausdorff measure \( (d\Sigma_q) \) can be explicitly written in terms of exterior product of differential forms as follows:

\[
(d\Sigma_q) = 2^{-m} |\Lambda|^{(m-n)} \prod_{i<j} (\lambda_i - \lambda_j)(H_i^*dH_i) (d\Lambda)
\]

where \( \Sigma_q = H_i^*\Lambda H_i \), is the nonsingular part of the spectral decomposition of \( \Sigma_q \), with \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \), \( \lambda_1 > \lambda_2 > \ldots > \lambda_n > 0 \) and \( H_i \in V_{n,m} \), the Stiefel manifold, i.e. the set of the real \( mxn \) full-rank matrices, such that \( H_i^*H_i = I_n \). Notice that \( (d\Lambda) \) is the exterior product of the positive eigenvalues of \( \Sigma_q \); see Uhlig (1994) and Diaz-Garcia and Gutierrez (1997).

**Remark:** As far as the differential form \( (H_i^*dH_i) \) is concerned, notice that it is invariant under the transformations \( H_i \rightarrow QH_i, [Q \in O(m)] \) and \( H_i \rightarrow H_iP, [P \in O(n)] \), where \( O(m) = V_{m,m} \) is the so-called orthogonal group. Therefore such differential form defines an (unnormalized) invariant measure on \( V_{n,m} \); see James (1954) and Muirhead (1982). This measure is invariant because for any region \( \mathcal{Z} \subset V_{n,m} \), \( \mu(Q\mathcal{Z}) = \mu(\mathcal{Z}) \), where \( \mu(\mathcal{Z}) = \int_H (H_i^*dH_i) \); so that \( \mu(\mathcal{Z}) \) represents the volume of the region \( \mathcal{Z} \) on the Stiefel manifold \( V_{n,m} \). The normalized invariant measure is a probability measure on \( V_{n,m} \) defined as

\[
\mu^*(\mathcal{Z}) = \frac{1}{Vol[V_{n,m}]} \int_{\mathcal{Z}} (H_i^*dH_i) = \frac{\mu(\mathcal{Z})}{Vol[V_{n,m}]} = \frac{2^n \pi^{mn/2}}{\Gamma_n\left(\frac{m}{2}\right)},
\]

with \( \Gamma_n(\cdot) \) denoting the multivariate Gamma function; see Muirhead (1982), pgg. 67-72.

The likelihood function of the normal singular population can be written as
where \( \Sigma^+ = H_i \Lambda^{-1} H_i \) is the Moore-Penrose generalized inverse of \( \Sigma \).

Thus, the posterior distribution of \( \Sigma \) is

\[
\pi(\Sigma | Y) = \frac{\pi(Y | \Sigma) \pi(\Sigma)}{\pi(Y)}
\]

(6) \[
\pi(\Sigma | Y) \propto l(\Sigma | Y) dP(\Sigma) (d\Sigma) \\
\propto \prod_{i=1}^{n} \lambda^{-T/2}_{i} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^+ Y'Y) \right\} d\Sigma
\]

which, according to the definition given by Diaz-Garcia and Gutierrez (2006), is a \( m \)-dimensional central, Singular Generalized Inverse Wishart of rank \( n \), with \( \nu = T - n + 1 \) degrees of freedom and scale matrix \( G = Y'Y \), denoted by \( W^+_m(n, \nu, G) \); that is we have that \( \Sigma'Y / Y \sim W^+_m(n, \nu, G) \).

We now expound the MCMC approach that will be subsequently used in the context of factor models, pretending that the posterior distribution of \( \Sigma \) is unknown and setting up a simple M-H algorithm for drawing from it. Therefore, the target density of the M-H algorithm is

\[
\pi(\Sigma) = l(\Sigma | Y) dP(\Sigma | Y) (d\Sigma)
\]

(7) \[
\pi(\Sigma) = l(\Sigma | Y) dP(\Sigma) (d\Sigma)
\]

and in what follows we develop a proposal distribution from which the candidate values for \( \Sigma \in \mathcal{S}^+_m(n) \) can be efficiently drawn. As outlined in the introduction, we explicitly take into account the curved geometry of the support of the target distribution, which prevents us from using standard candidate-generating densities such as the random walk M-H chain; see, e.g., Chib (2001). In fact, the crucial issue in the design of the M-H step is that a linear combination of any two elements of the manifold \( \mathcal{S}^+_m(n) \) may not be an element of the same manifold. We resolve this issue by setting up a proposal distribution which is based on a mixture of Wishart Singular distributions. Yet, prior presenting such proposal distribution we briefly define the Wishart Singular distribution. The following definition is taken from Diaz-Garcia et al. (1997) and adapted to our objective. If the matrix \( W \) has a \( m \)-dimensional Wishart Singular distribution with degrees of freedom \( \nu \) and scale matrix \( C \), where both \( W \)
and $C$ are $m \times m$ symmetric positive semi-definite matrices of rank $n < m$ and $v \geq n$, then the density of $W$ is given by

$$
\begin{align*}
\frac{1}{2^{m/2} \Gamma_n(\frac{v}{2})} & |L|^{-v/2} |A|^{(v-m)/2} \exp\left(-\frac{1}{2} \text{tr}(C^{-1}W)\right) (dW)^{+} \\
& = f_W(W)(dW)^{+}
\end{align*}
$$

and this fact is denoted by writing $W \sim SW_m(n, \nu, C)$, where $W = H_1 \Lambda H_1'$ and $C = U_1L U_1'$ are the nonsingular part of the spectral decomposition of $W$ and $C$ respectively, with $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $L = \text{diag}(l_1, l_2, \ldots, l_n)$ such that $\lambda_i, i = 1, \ldots, n$ and $l_i, i = 1, \ldots, n$ are the $n$ distinct positive eigenvalues (in decreasing order) of $W$ and $C$ respectively, and $H_1, U_1 \in V_{n,m}$; $C^{-1}$ represents any generalized inverse of $C$ such that $CC^{-1}C = C$ and we assume $C^{-1} = U_1L^{-1}U_1'$. Finally, $(dW)^{+}$ is the Hausdorff measure defined on $S_m^{+}(n)$, which can be written in terms of exterior product of differential forms as follows:

$$
(dW)^{+} = 2^{-n} |\Lambda|^{(m-n)} \prod_{i<j} (\lambda_i - \lambda_j) (H_1 dH_1) (d\Lambda).
$$

We note that, like for the non singular Wishart distribution, $E[W] = \nu C$. Moreover, we show in the Appendix how we can straightforwardly draw from the Wishart Singular distribution with integer degrees of freedom.

We are now in a position to present the mixture of Wishart Singular distributions we adopt as proposal distribution. Given the current value $\Sigma^{(0)}_i$, we suggest to draw the candidate value $\Sigma^{(1)}_i$ from the following proposal density:

$$
\begin{align*}
\frac{dQ(\Sigma^{(0)}_i, \Sigma^{(1)}_i)}{d\Sigma^{(1)}_i} & = \int dQ(\Sigma^{(0)}_i, \Sigma^{(1)}_i / A) dQ(A) \\
& = \int dQ(\Sigma^{(0)}_i, \Sigma^{(1)}_i / A) dQ(A)
\end{align*}
$$

where $Q(A)$ is the distribution, to be defined below, of the matrix $A$, with $A \in S_m^{+}$, the open cone of the $m \times m$ symmetric positive definite matrices, and
$$Q(\Sigma_0^{(0)}, \Sigma_0^{(1)}/A) \equiv SW_m(n,v,v^{-1}A\Sigma_0^{(0)}A)$$ \text{ with } \quad E\left[\Sigma_0^{(1)}/\Sigma_0^{(0)},A\right] = A\Sigma_0^{(0)}A. \quad \text{It is worth noting that the scale matrix } C = v^{-1}A\Sigma_0^{(0)}A \text{ depends on the current value of the chain, } \Sigma_0^{(0)}, \text{ and it is a } m \times m \text{ symmetric positive semi-definite matrix of rank } n \text{ with probability one.}

As regards the distribution for } A, \text{ we find it convenient to choose the following mixture of a degenerate distribution, } Q_1(A), \text{ and a Wishart distribution, } Q_2(A):

\begin{equation}
Q(A) = \lambda \cdot Q_1(A) + (1 - \lambda) \cdot Q_2(A)
\end{equation}

where } Q_1(A) = 1 \text{ for } A = I_m \text{ and zero elsewhere, } Q_2(A) = W_m(v_2, v_2^{-1}I_m) \text{ with } v_2 > m^8 \text{ and } 0 \leq \lambda \leq 1. \text{ By adopting this distribution for } A, \text{ we obtain that the mean of } Q(A) \text{ is equal to the identity matrix. Further, the proposal density } dQ(\Sigma_0^{(0)}, \Sigma_0^{(1)}) \text{ can be written as }

\begin{equation}
dQ(\Sigma_0^{(0)}, \Sigma_0^{(1)}) = \lambda \cdot dQ(\Sigma_0^{(0)}, \Sigma_0^{(1)}/I_m) + (1 - \lambda) \int_{S_m^+} dQ(\Sigma_0^{(0)}, \Sigma_0^{(1)}/A) dQ_2(A)
\end{equation}

where } Q(\Sigma_0^{(0)}, \Sigma_0^{(1)}/I_m) \equiv SW_m(n,v,v^{-1}\Sigma_0^{(0)}). \text{ In order to understand the rationale of our specification, firstly we consider what happens if } \lambda = 1. \text{ In this case the proposal density reduces to } dQ(\Sigma_0^{(0)}, \Sigma_0^{(1)}) = dQ(\Sigma_0^{(0)}, \Sigma_0^{(1)}/I_m), \text{ so that } \Sigma_0^{(1)}/\Sigma_0^{(0)} \sim SW_m(n,v,v^{-1}\Sigma_0^{(0)}) \text{ with } E\left[\Sigma_0^{(1)}/\Sigma_0^{(0)}\right] = \Sigma_0^{(0)}. \text{ Therefore, if } \lambda = 1 \text{ the candidate-generating density is a } \textit{random walk} \text{ M-H chain defined on the manifold } S_m^+(n). \text{ In this way we obtain that: i) the proposal takes explicitly into account the curved geometry of the original parameter space; ii) the proposal is centered on the current value of the chain, } \Sigma_0^{(0)}. \text{ Although the proposal distribution } Q(\Sigma_0^{(0)}, \Sigma_0^{(1)}) \equiv SW_m(v,v^{-1}\Sigma_0^{(0)}) \text{ seems the natural choice for the problem in hand, a crucial issue prevents its use. In fact, as noted by}

\footnote{Following Press (1982), we denote by } W_m(v,S) \text{ the } m \text{-dimensional Wishart distribution with } v \text{ degrees of freedom and scale matrix } S.
Bhimasankaram and Sengupta (1991), pg. 474, if \( \mathbf{W} \sim SW_m(\nu, \mathbf{C}) \) then the column space of the matrices \( \mathbf{W} \) and \( \mathbf{C} \) are the same with probability one. This property of the Singular Wishart distribution has remarkable implications for the mixing of the M-H chain: the starting value chosen for \( \Sigma_\eta \) and all the singular covariance matrices subsequently drawn will have their column spaces in common. As a result, the Markov chain will be trapped in a subset of \( S^+_m(n) \) which is determined by the choice of the starting value.

If \( \lambda = 0 \), then \( dQ(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta) = \int dQ(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / \mathbf{A})dQ_2(\mathbf{A}) \). This proposal density is a mixture of the Wishart Singular distributions \( SW_m(n, \nu, \nu^{-1} \mathbf{A}\Sigma^{(0)}_\eta \mathbf{A}) \), for \( \mathbf{A} \in S_m \), each one entering the mixture with weight given by the Wishart density of \( \mathbf{A} \), \( q_2(\mathbf{A}) \).

Moreover, since the Wishart density \( q_2(\mathbf{A}) \) is centered on the identity matrix, the mixture distribution assigns the highest weight to the Wishart Singular distribution \( SW_m(n, \nu, \nu^{-1} \Sigma^{(0)}_\eta) \). It is worth noticing that by averaging over all the possible values of \( \mathbf{A} \), actually we are averaging the Wishart Singular distributions over all the possible values of the scale matrix \( \mathbf{C} = \nu^{-1} \mathbf{A}\Sigma^{(0)}_\eta \mathbf{A} \). We obtain in this way that the M-H chain moves through the whole parameter space \( S^+_m(n) \).

Finally, if \( 0 < \lambda < 1 \) we draw \( \Sigma^{(1)}_\eta \) from \( SW_m(n, \nu, \nu^{-1} \Sigma^{(0)}_\eta) \) with probability equal to \( \lambda \). This specification implies both a positive probability of drawing a singular covariance matrix whose column space is the same of \( \Sigma^{(0)}_\eta \), and a positive probability of moving toward a different column space. In theory, at the start of the chain a low value of \( \lambda \) seems to be preferable in order to improve the mixing of the M-H chain, whereas the opposite is true after convergence. In practice, the best results have been obtained using the proposal density given in eq. (12) with \( \lambda = 0.7 \).

For drawing \( \Sigma^{(1)}_\eta \) from the proposal density given in eq. (12), one further issue has to be addressed for. That is, we must calculate \( \int dQ(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / \mathbf{A})dQ_2(\mathbf{A}) \) or, at least, be able to draw from it. It is interesting to notice that we can avoid such computation by the following trick. We augment the model, including also the matrix \( \mathbf{A} \) among the
unknown parameters, and we specify the proposal density for the transition from \( \{\Sigma^{(0)}_\eta, A^{(0)}\} \) to \( \{\Sigma^{(1)}_\eta, A^{(1)}\} \) as

\[
(13) \quad dQ(\{\Sigma^{(0)}_\eta, A^{(0)}\}, \{\Sigma^{(1)}_\eta, A^{(1)}\}) = dQ(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / A^{(1)}) \cdot dQ(A^{(1)})
\]

where \( Q(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / A^{(1)}) \equiv SW_m(n, v, v^{-1}A^{(1)}\Sigma^{(0)}_\eta A^{(1)}) \) and \( Q(A) \) is defined as in (11).

The proposal density in (13) implies that, given \( \{\Sigma^{(0)}_\eta, A^{(0)}\} \), we firstly draw \( A^{(1)} \) from \( Q(A) \) (independently from the values \( \{\Sigma^{(0)}_\eta, A^{(0)}\} \)), and then, conditional on \( A^{(1)} \) (and \( \Sigma^{(0)}_\eta \)), we draw \( \Sigma^{(1)}_\eta \) from \( SW_m(n, v, v^{-1}A^{(1)}\Sigma^{(0)}_\eta A^{(1)}) \).

Since the likelihood function does not depend on \( A \) and we further assume that \( A \) and \( \Sigma_\eta \) are a priori independent, the posterior density \( dP(\Sigma_\eta, A / Y) \) will be proportional to the target density \( \pi(\Sigma_\eta, A) = dP(A) \cdot dP(\Sigma_\eta) / l(\Sigma_\eta, Y) (d\Sigma_\eta)(dA) = dP(A) \cdot \pi(\Sigma_\eta) \),

where \( dP(A) \) denotes the prior density of \( A \) and \( \pi(\Sigma_\eta) \) is defined as in (7).

Therefore, it is worth noting that after convergence the Markov chain values of \( \Sigma_\eta \) represent a sample from \( \pi(\Sigma_\eta) \), i.e. from the posterior density of the original model.

Moreover, we find it convenient to assume that \( dP(A) = dQ(A) \), since this assumption implies a simplification in the calculation of the acceptance probability.

The M-H algorithm is completed by accepting the candidate values \( \{\Sigma^{(1)}_\eta, A^{(1)}\} \) drawn from (13) with probability

\[
\alpha(\{\Sigma^{(0)}_\eta, A^{(0)}\}, \{\Sigma^{(1)}_\eta, A^{(1)}\}) =
\]

\[
(14) \quad \begin{cases} 
\min \left\{ \frac{\pi(\{\Sigma^{(1)}_\eta, A^{(1)}\}) \cdot dQ(\{\Sigma^{(1)}_\eta, A^{(1)}\}, \{\Sigma^{(0)}_\eta, A^{(0)}\})}{\pi(\{\Sigma^{(0)}_\eta, A^{(0)}\}) \cdot dQ(\{\Sigma^{(0)}_\eta, A^{(0)}\}, \{\Sigma^{(1)}_\eta, A^{(1)}\})}, 1 \right\} \\
\quad \text{if } \pi(\{\Sigma^{(0)}_\eta, A^{(0)}\}) \cdot dQ(\{\Sigma^{(0)}_\eta, A^{(0)}\}, \{\Sigma^{(1)}_\eta, A^{(1)}\}) > 0; \\
1 \text{ otherwise}
\end{cases}
\]

13
where

\[
\frac{\pi\left(\{\Sigma^{(1)}_\eta, A^{(1)}_\eta\}\right) dQ\left(\{\Sigma^{(1)}_\eta, A^{(1)}_\eta\}, \{\Sigma^{(0)}_\eta, A^{(0)}_\eta\}\right)}{\pi\left(\{\Sigma^{(0)}_\eta, A^{(0)}_\eta\}\right) dQ\left(\{\Sigma^{(0)}_\eta, A^{(0)}_\eta\}, \{\Sigma^{(1)}_\eta, A^{(1)}_\eta\}\right)} = \frac{\pi\left(\Sigma^{(1)}_\eta\right) dP\left(A^{(1)}_\eta\right) dQ\left(A^{(0)}_\eta\right) dQ\left(\Sigma^{(1)}_\eta, \Sigma^{(0)}_\eta / A^{(0)}_\eta\right)}{\pi\left(\Sigma^{(0)}_\eta\right) dP\left(A^{(0)}_\eta\right) dQ\left(A^{(0)}_\eta\right) dQ\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / A^{(1)}_\eta\right)}
\]

\[
= \frac{\pi\left(\Sigma^{(1)}_\eta\right) dQ\left(\Sigma^{(1)}_\eta, \Sigma^{(0)}_\eta / A^{(0)}_\eta\right)}{\pi\left(\Sigma^{(0)}_\eta\right) dQ\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / A^{(1)}_\eta\right)}
\]

with 

\[Q\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta / A^{(1)}_\eta\right) = SW_{\eta}\left(t, \nu, \nu^{-1} A^{(1)}_\eta \Sigma^{(0)}_\eta A^{(1)}_\eta\right)\]

and

\[Q\left(\Sigma^{(1)}_\eta, \Sigma^{(0)}_\eta / A^{(0)}_\eta\right) = SW_{\eta}\left(t, \nu, \nu^{-1} A^{(0)}_\eta \Sigma^{(1)}_\eta A^{(0)}_\eta\right)\].

Moreover, by denoting the non-null eigenvalues of \(\Sigma^{(0)}_\eta\) and \(\Sigma^{(1)}_\eta\) as \(\lambda_{i(0)}\) and \(\lambda_{i(1)}\), \(i = 1, \ldots, n\), respectively, the ratio above can be written as:

\[
\prod_{i=1}^{n} \lambda_{i(1)}^{(T + 2m - n + 1)/2} \exp\left\{-\frac{1}{2} \text{tr}\left(\Sigma^{(0)}_\eta \Sigma^{(0)}_\eta - \Sigma^{(1)}_\eta \Sigma^{(1)}_\eta\right)\right\} \left(d\Sigma^{(1)}_\eta\right) \left(d\Sigma^{(0)}_\eta\right)^* \frac{q\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta, A^{(0)}_\eta\right) (d\Sigma^{(1)}_\eta)^*}{q\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta, A^{(0)}_\eta\right) (d\Sigma^{(0)}_\eta)^*}
\]

where \(d\Sigma^{(1)}_\eta = q\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta, A^{(0)}_\eta\right) (d\Sigma^{(1)}_\eta)^*\) and \(d\Sigma^{(0)}_\eta = q\left(\Sigma^{(1)}_\eta, \Sigma^{(0)}_\eta, A^{(1)}_\eta\right) (d\Sigma^{(0)}_\eta)^*\).

Since \((d\Sigma^{(0)}_\eta)^* = \frac{2^{-n}}{2^{-m}} (d\Sigma^{(0)}_\eta)\), the two Hausdorff measures are proportional for \(n\) and \(m\) fixed, and eq. (15) simplifies to

\[
\prod_{i=1}^{n} \lambda_{i(1)}^{(T + 2m - n + 1)/2} \exp\left\{-\frac{1}{2} \text{tr}\left(\Sigma^{(0)}_\eta \Sigma^{(0)}_\eta - \Sigma^{(1)}_\eta \Sigma^{(1)}_\eta\right)\right\} q\left(\Sigma^{(0)}_\eta, \Sigma^{(1)}_\eta, A^{(0)}_\eta\right)
\]

Therefore, we don’t need to take into account the Hausdorff measure in the calculation of the M-H acceptance probability, as well as when drawing from the proposal density (13). Finally, it is worth noting that both the proposal distributions (12) and (13) depend not only on \(\lambda\) but also on the degrees of freedom \(\nu_2\) and \(\nu\). By
fine tuning these setting parameters so that a M-H rejection rate in the range 75-85% is obtained, we succeed in efficiently moving the chain through the whole support of the target density.

2.1 Bayesian inference on the singular covariance matrix

TO BE WRITTEN

3. Bayesian factor analysis

4. A multivariate local level model with common trends

5. Conclusion
APPENDIX: simulation of a Wishart Singular distribution with integer degrees of freedom

If the matrix $W$ has an $m$-dimensional Wishart Singular distribution with integer degrees of freedom $v \geq n$ and scale matrix $C$, where both $W$ and $C$ are $m \times m$ symmetric positive semi-definite, singular matrices of rank $n < m$ then we can draw $W$ from $SW_m(v, C)$ as follows:

1) for $C = U \Sigma U^T$, where $U \Sigma U^T$ is the nonsingular part of the spectral decomposition of $C$, calculate $B = U \Sigma^{1/2}$;

2) generate $x_1, \ldots, x_v$ independently from $N(0, I_n)$;

3) calculate $Z_i = Bx_i$, $i = 1, \ldots, v$;

4) then $W = \sum_{i=1}^{v} Z_i Z_i^T$ is a draw from $SW_m(v, C)$. 


References


