Central European Journal of Economic Modelling and Econometrics

A Note on Compatible Prior Distributions in Univariate Finite Mixture and Markov-Switching Models

Łukasz Kwiatkowski*

Submitted: 1.08.2015, Accepted: 13.10.2015

Abstract

Finite mixture and Markov-switching models generalize and, therefore, nest specifications featuring only one component. While specifying priors in the general (mixture) model and its special (single-component) case, it may be desirable to ensure that the prior assumptions introduced into both structures are compatible in the sense that the prior distribution in the nested model amounts to the conditional prior in the mixture model under relevant parametric restriction. The study provides the rudiments of setting compatible priors in Bayesian univariate finite mixture and Markov-switching models. Once some primary results are delivered, we derive specific conditions for compatibility in the case of three types of continuous priors commonly engaged in Bayesian modeling: the normal, inverse gamma, and gamma distributions. Further, we study the consequences of introducing additional constraints into the mixture model's prior on the conditions. Finally, the methodology is illustrated through a discussion of setting compatible priors for Markov-switching AR(2) models.

Keywords: Bayesian inference, prior coherence, prior compatibility, exponential family

JEL Classification: C11

*Cracow University of Economics; e-mail: kwiatkol@uek.krakow.pl

1 Introduction

Consider two statistical models, say, M_G (the general one) and M_R (the restricted one), such that the latter constitutes a special case of the former under some parametric restriction, and let vectors $\theta^{(G)}$ and $\theta^{(R)}$ collect their parameters, respectively. Note that $\theta^{(G)}$ includes $\theta^{(R)}$, which thereby is the vector of common parameters (as opposed to the vector ofM_G's specific coefficients, say, γ, so that $\theta^{(G)} = (\theta^{(R)\prime}\ \gamma')'$, with A' symbolizing the transpose of any matrix A). Let γ_0 be the value of γ under which M_G collapses to M_R. In what follows, we adopt notational convention under which, generally, $\pi_{\underline{\omega}}(\omega|M)$ denotes the p.d.f. of some random variable $\underline{\omega}$ at $\underline{\omega} = \omega$ under model M. Analogously, $\pi_{\underline{\omega}|\underline{\gamma}}(\omega|\gamma, M)$ stands for the p.d.f. of $\underline{\omega}$'s conditional distribution at $\underline{\omega} = \omega$ given $\underline{\gamma} = \gamma$. Finally, to avoid measure-theoretic intricacies, though with some abuse of notation, we use the above symbols of density functions to refer to the underlying distributions as well.
Within a non-Bayesian statistical framework, that M_G nests M_R amounts to the equality of corresponding sample distributions under the nesting constraint, i.e., $\pi_{\underline{y}|\underline{\theta^{(R)}}}(y|\theta^{(R)}, M_R) = \pi_{\underline{y}|\underline{\theta^{(R)}},\underline{\gamma}}(y|\theta^{(R)}, \gamma = \gamma_0, M_G)$ for any $y \in Y \subseteq \mathbb{R}^T$. However, should the models in question be regarded Bayesian, then nesting M_R in M_G would also require that the prior information introduced in the former be "nested" in the one incorporated into the general structure. It follows then that $\pi_{\underline{\theta^{(R)}}}(\theta^{(R)}|M_R)$ should be induced from $\pi_{\underline{\theta^{(G)}}}(\theta^{(G)}|M_G)$ via conditioning upon the reducing restriction. The definition below formalizes the concept of such prior compatibility.

Definition 1 (Prior compatibility)**.** If the prior distributions: $\pi_{\underline{\theta^{(R)}}}(\theta^{(R)}|M_R)$ and $\pi_{\underline{\theta^{(G)}}}(\theta^{(G)}|M_G)$, satisfy the condition:

$$\pi_{\underline{\theta^{(R)}}}\left(\theta^{(R)}|M_R\right) = \pi_{\underline{\theta^{(R)}}|\underline{\gamma}}\left(\theta^{(R)}|\gamma = \gamma_0, M_G\right), \tag{1}$$

then they are called compatible, and the models M_G and M_R are said to feature compatible prior structures.

Note that if $\theta^{(R)}$ and γ in the M_G model are *a priori* independent, then it is required for the prior compatibility that the prior of $\theta^{(R)}$ be the same in both models, i.e., $\pi_{\underline{\theta^{(R)}}}(\theta^{(R)}|M_R) = \pi_{\underline{\theta^{(R)}}}(\theta^{(R)}|M_G)$.
The idea of specifying compatible (or, coherent) prior distributions has been originated by Dickey (1974) and Poirier (1985) in the context of hypothesis testing within linear models. We refer the reader to Consonni and Veronese (2008) for a recent study and literature review on various forms of prior compatibility across linear models.
Obviously, the idea of establishing compatible prior structures over various models does not pertain to the class of the linear specifications solely, but applies whenever the nesting comes into play. In particular, the mixture (and Markov-switching) models nest their single-component counterparts, the latter being derived from

the former via relevant equality restrictions. Perversely, one may argue, however, that there is no compelling reason within the subjective framework to relate priors across models, since they express subjective opinions conditionally on a different state of information. Nevertheless, ensuring prior compatibility across various model specifications appears crucial to the model comparison (usually performed via recognizably prior-sensitive Bayes factors), for reconciling the models' prior structures sheds some layer of arbitrariness (Dawid and Lauritzen 2001, Consonni and Veronese 2008). In particular, within the finite mixture and Markov-switching class of models, specifying compatible priors may be desirable for testing the relevance of incorporating the mixture (switching) structure into the otherwise single-component specification. To the author's best knowledge, the issue of prior compatibility within the mixture models has not been raised in the literature so far. Therefore, in the present research, we take an interest in settling compatible prior structures for the mixture models (and the Markov-switching structures alike) and their single-component counterparts.

In Section 2 we lay the basic foundations of establishing compatible prior structures within the finite mixture and the Markov-switching model frameworks, and arrive at the basic lemma. The results incline us to focus next on exponential families of prior distributions, for three representatives of which, namely the normal, inverse gamma, and gamma distributions, we derive in 3 explicit conditions relating the hyperparameters of the general and the nested model. Section 4 is devoted to the cases in which the priors are subject to certain restrictions, such as the ones enforcing identifiability of the mixture components (via an inequality constraint imposed on a group of mixture parameters) or some sort of regularity (e.g., the second-order stationarity). In Section 5, the methodology is illustrated with a discussion of setting compatible priors for a class of Markov-switching AR(2) models. Finally, we conclude with several remarks indicating a need and clear directions for further research on compatible priors for mixture and switching models.

2 Prior compatibility in the mixture and Markov-switching models

Consider a single-component model, M_1, with parameters collected in

$$\theta^{(1)} = (\delta' \; \lambda_{1,1} \; \lambda_{1,2} \; \dots \; \lambda_{1,n})' \in \Theta^{(1)}, \qquad n \in \mathbb{N},$$

and the general, K-component mixture model, M_K, $K \in \mathbb{N}$, with parameters

$$\theta^{(K)} = \left(\delta' \; \lambda_1^{(K)\prime} \; \lambda_2^{(K)\prime} \; \dots \; \lambda_n^{(K)\prime} \; \eta'\right)' \in \Theta^{(K)}.$$

The following remarks clarify our notational convetion:

1. The vector δ is comprised of the parameters that are non-mixture and common to both models.

2. The parameters $\lambda_{1,j}$ $(j = 1, 2, \ldots, n)$ in the model M_1 are scalar, with the first subscript indicating that the model features a single component.

3. Each vector $\lambda_j^{(K)}$ $(j = 1, 2, \ldots, n)$ in the model M_K collects K parameters that arise as a result of introducing the K-component mixture structure into the corresponding (scalar) parameter $\lambda_{1,j}$ in M_1, so that $\lambda_j^{(K)} = (\lambda_{1,j}\ \lambda_{2,j}\ \ldots\ \lambda_{K,j})'$. Note that the first coordinate in $\lambda_j^{(K)}$, denoted by $\lambda_{1,j}$, coincides with the corresponding parameter in the single-component model.

4. The vector $\eta = (\eta_1\ \eta_2\ \ldots\ \eta_K)'$ in the model M_K contains the probability parameters:

   (a) If M_K is a finite mixture model, then η_i $(i = 1, 2, \ldots, K)$ are the mixture probabilities, and $\eta \in \Delta^{(K-1)}$, where $\Delta^{(K-1)}$ denotes the unit (K-1)-simplex.

   (b) If M_K is a Markov-switching model, with $\{S_t;\ t = 0, 1, \ldots\}$ forming the underlying K-state (homogenous) Markov chain, then η_i $(i = 1, 2, \ldots, K)$ are the rows of transition matrix $P = [\eta_{ij}]_{i,j=1,2,\ldots,K}$, $\eta_{ij} \equiv \Pr(S_t = j|S_{t-1} = i)$, i.e., $\eta_i = (\eta_{i1}\ \eta_{i2}\ \ldots\ \eta_{iK}) \in \Delta^{(K-1)}$, and therefore $\eta \in (\Delta^{(K-1)})^K$. For simplicity, though without loss of generality, we assume that the chain's initial state distribution: $\xi = (\xi_1\ \xi_2\ \ldots\ \xi_K)'$, $\xi_i \equiv \Pr(S_0 = i)$, $i = 1, 2, \ldots, K$, is either known (e.g., a uniform distribution) or equal to the chain's ergodic distribution (which introduces into the probabilities ξ_i's conditioning upon the transition matrix, $\xi_i \equiv \Pr(S_0 = i|P)$).

Notice that the two: M_1 and M_K, represent the extremes, i.e, at one end, there is the single-component model M_1, whereas at the other – the specification M_K, in which all $\lambda_j^{(K)}$'s constitute the mixture counterparts of $\lambda_{1,j}$'s in M_1. Obviously, there are $2^n - 2$ specifications in between, such that only some of $\lambda_j^{(K)}$'s are actually the vectors of mixture parameters, whereas the other ones remain equivalent to the corresponding coefficients in the single-component model. These "intermediate" model structures encompass M_1 on the one hand, and, on the other, are nested within the most general one, i.e., M_K. Nevertheless, we limit most of our further considerations only to the two extreme cases, for the reason that, under the assumptions of our analysis, establishing compatible priors for the two: the single-component model and any of the "intermediate" constructions, comes down to the same framework by means of relegating those $\lambda_{1,j}$'s that are non-mixture in both specifications to the vector of the common parameters, δ. In a similar fashion, compatibility of the "intermediate" and the general model can be settled, which would require including also the probabilities η in the common parameters vector. We revisit the issue in the final paragraph of Section 5.

In what follows, for both models in question, prior independence is assumed between the parameter vector's components:

$$\pi_{\underline{\theta^{(1)}}}(\theta^{(1)}|M_1) = \pi_{\underline{\delta}}(\delta|M_1)\prod_{j=1}^{n}\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1)\,, \tag{2}$$

$$\pi_{\underline{\theta^{(K)}}}(\theta^{(K)}|M_K) = \pi_{\underline{\delta}}(\delta|M_K)\left[\prod_{j=1}^{n}\pi_{\underline{\lambda_j^{(K)}}}(\lambda_j^{(K)}|M_K)\right]\pi_{\underline{\eta}}(\eta|M_K). \tag{3}$$

Moreover, for each $j = 1, 2, \ldots, n$, also the individual coordinates of $\lambda_j^{(K)}$ are presumed *a priori* independent:

$$\pi_{\underline{\lambda_j^{(K)}}}(\lambda_j^{(K)}|M_K) = \prod_{i=1}^{K}\pi_{\underline{\lambda_{i,j}}}(\lambda_{i,j}|M_K)\,. \tag{4}$$

Finally, all the priors under consideration are assumed proper, for it may be shown that setting improper priors in mixture models yields improper posteriors (see Roeder and Wasserman 1997, Frühwirth-Schnatter 2006).

Resting upon (2) and (3), the priors: $\pi_{\underline{\theta^{(1)}}}(\theta^{(1)}|M_1)$ and $\pi_{\underline{\theta^{(K)}}}(\theta^{(K)}|M_K)$, are compatible if the two conditions are met simultaneously:

$$\pi_{\underline{\delta}}\,(\delta|M_1) = \pi_{\underline{\delta}}\,(\delta|M_K) \tag{5}$$

and, for all $j = 1, 2, \ldots, n$,

$$\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1) \propto \pi_{\underline{\lambda_j^{(K)}}}\left(\lambda_j^{(K)}|\lambda_{1,j} \equiv \lambda_{2,j} = \lambda_{3,j} = \ldots = \lambda_{K,j}, M_K\right). \tag{6}$$

Note that the postulated conditions do not explicitly concern the mixture probabilities (η), as these are either entirely absent from the reduced model (M_1) or contained in the vector δ (in the case of establishing compatible prior structures for the general and some "intermediate" model specification; then, (5) applies).

In order to rewrite (6) in terms of Definition 1, the general model needs to be suitably reparametrized. Let $\widetilde{M}_K$ be the reparametrized model, with the parameters grouped in

$$\widetilde{\theta}^{(K)} = \left(\delta'\ \widetilde{\lambda}_1^{(K)\prime}\ \widetilde{\lambda}_2^{(K)\prime}\ \ldots\ \widetilde{\lambda}_n^{(K)\prime}\ \eta'\right)' \in \widetilde{\Theta}^{(K)}.$$

Each $\widetilde{\lambda}_j^{(K)}$ $(j = 1, 2, \ldots, n)$ is obtained from the corresponding $\lambda_j^{(K)}$ via a transformation $g : \mathbb{R}^K \to \mathbb{R}^K$:

$$\widetilde{\lambda}_j^{(K)} = g(\lambda_j^{(K)}) = \begin{pmatrix} g_1(\lambda_{1,j}) \\ g_2(\lambda_{2,j}) \\ g_3(\lambda_{3,j}) \\ \vdots \\ g_K(\lambda_{K,j}) \end{pmatrix} = \begin{pmatrix} \lambda_{1,j} \\ \lambda_{2,j} - \lambda_{1,j} \\ \lambda_{3,j} - \lambda_{1,j} \\ \vdots \\ \lambda_{K,j} - \lambda_{1,j} \end{pmatrix} \equiv \begin{pmatrix} \lambda_{1,j} \\ \tau_j \end{pmatrix},$$

with $\tau_j = (\tau_{2,j} \quad \tau_{3,j} \quad \ldots \quad \tau_{K,j})'$ collecting the contrasts $\tau_{i,j} = \lambda_{i,j} - \lambda_{1,j}$ $(i = 2, 3, \ldots, K)$. The inverse transformation follows as

$$g^{-1}(\widetilde{\lambda}_j^{(K)}) = \begin{pmatrix} g_1^{-1}(\lambda_{1,j}) \\ g_2^{-1}(\tau_{2,j}) \\ g_3^{-1}(\tau_{3,j}) \\ \vdots \\ g_K^{-1}(\tau_{K,j}) \end{pmatrix} = \begin{pmatrix} \lambda_{1,j} \\ \tau_{2,j} + \lambda_{1,j} \\ \tau_{3,j} + \lambda_{1,j} \\ \vdots \\ \tau_{K,j} + \lambda_{1,j} \end{pmatrix} = \begin{pmatrix} \lambda_{1,j} \\ \lambda_{2,j} \\ \lambda_{3,j} \\ \vdots \\ \lambda_{K,j} \end{pmatrix} \equiv \lambda_j^{(K)}.$$

Owing to the fact that $\left| \dfrac{\partial g^{-1}\left(\widetilde{\lambda}_j^{(K)}\right)}{\partial \widetilde{\lambda}_j^{(K)}} \right| = 1$, the p.d.f. of $\widetilde{\lambda}_j^{(K)}$'s prior can be easily derived:

$$\begin{aligned} \pi_{\underline{\widetilde{\lambda}_j^{(K)}}}(\widetilde{\lambda}_j^{(K)}|\widetilde{M}_K) &= \pi_{\underline{\lambda_j^{(K)}}}\left(g^{-1}(\widetilde{\lambda}_j^{(K)})|M_K\right) = \\ &= \pi_{\underline{\lambda_{1,j}}}\left(g_1^{-1}(\lambda_{1,j})|M_K\right)\prod_{i=2}^{K}\pi_{\underline{\lambda_{i,j}}}\left(g_i^{-1}(\tau_{i,j})|M_K\right) = \\ &= \pi_{\underline{\lambda_{1,j}}}\left(\lambda_{1,j}|M_K\right)\prod_{i=2}^{K}\pi_{\underline{\lambda_{i,j}}}\left(\tau_{i,j} + \lambda_{1,j}|M_K\right). \end{aligned} \tag{7}$$

Now, recall that M_1 results from M_K under the equality constraint of all coordinates within each vector $\lambda_j^{(K)}$, i.e.,

$$\lambda_{1,j} = \lambda_{2,j} = \cdots = \lambda_{K,j}, \qquad j = 1, 2, \ldots n. \tag{8}$$

In the model $\widetilde{M}_K$, (8) is equivalent to setting all the corresponding contrasts to zero:

$$\lambda_{1,j} = \lambda_{2,j} = \cdots = \lambda_{K,j} \Leftrightarrow \tau_{2,j} = \tau_{3,j} = \cdots = \tau_{K,j} = 0 \Leftrightarrow \tau_j = 0_{[(K-1)\times 1]}.$$

Conditions (5) and (6) can now be restated in terms of the reparametrized model, which leads us to formulate Lemma 2, summarising the above analysis.

Lemma 2. *Assume that:*

1. *In the single-component model, M_1, parameters in δ are* a priori *independent of all the others:*

$$\pi_{\underline{\delta},\underline{\lambda_{1,1}},\underline{\lambda_{1,2}},\ldots,\underline{\lambda_{1,n}}}\left(\delta,\lambda_{1,1},\lambda_{1,2},\ldots,\lambda_{1,n}|M_1\right)= \\ =\pi_{\underline{\delta}}(\delta|M_1)\pi_{\underline{\lambda_{1,1}},\underline{\lambda_{1,2}},\ldots,\underline{\lambda_{1,n}}}\left(\lambda_{1,1},\lambda_{1,2},\ldots,\lambda_{1,n}|M_1\right).$$

2. *In the reparametrized mixture model, $\widetilde{M}_K$, parameters in δ are* a priori *independent of all the others:*

$$\pi_{\underline{\delta},\underline{\lambda_{1,1}},\underline{\tau_1},\underline{\lambda_{1,2}},\underline{\tau_2},\ldots,\underline{\lambda_{1,n}},\underline{\tau_n}}\left(\delta,\lambda_{1,1},\tau_1,\lambda_{1,2},\tau_2,\ldots,\lambda_{1,n},\tau_n|\widetilde{M}_K\right)= \\ =\pi_{\underline{\delta}}\left(\delta|\widetilde{M}_K\right)\pi_{\underline{\lambda_{1,1}},\underline{\tau_1},\underline{\lambda_{1,2}},\underline{\tau_2},\ldots,\underline{\lambda_{1,n}},\underline{\tau_n}}\left(\lambda_{1,1},\tau_1,\lambda_{1,2},\tau_2,\ldots,\lambda_{1,n},\tau_n|\widetilde{M}_K\right).$$

3. *All prior distributions under consideration are proper.*

Then, distributions $\pi_{\underline{\theta^{(1)}}}\left(\theta^{(1)}|M_1\right)$ and $\pi_{\underline{\theta^{(K)}}}\left(\theta^{(K)}|M_K\right)$ - likewise, $\pi_{\underline{\widetilde{\theta}^{(K)}}}\left(\widetilde{\theta}^{(K)}|\widetilde{M}_K\right)$ - are compatible, i.e.

$$\pi_{\underline{\delta},\underline{\lambda_{1,1}},\underline{\lambda_{1,2}},\ldots,\underline{\lambda_{1,n}}}\left(\delta,\lambda_{1,1},\lambda_{1,2},\ldots,\lambda_{1,n}|M_1\right)= \\ =\pi_{\underline{\delta},\underline{\lambda_{1,1}},\underline{\lambda_{1,2}},\ldots,\underline{\lambda_{1,n}}|\underline{\tau_1},\underline{\tau_2},\ldots,\underline{\tau_n}}\left(\delta,\lambda_{1,1},\lambda_{1,2},\ldots,\lambda_{1,n}|\tau_1=\ldots=\tau_n=0_{[(K-1)\times 1]},\widetilde{M}_K\right), \tag{9}$$

iff the following conditions are satisfied:

$$\pi_{\underline{\delta}}(\delta|M_1)=\pi_{\underline{\delta}}(\delta|\widetilde{M}_K), \tag{10}$$

and, for each $j=1,2,\ldots,n$,

$$\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1)=\pi_{\underline{\lambda_{1,j}}|\underline{\tau_j}}\left(\lambda_{1,j}|\tau_j=0_{[(K-1)\times 1]},\widetilde{M}_K\right). \tag{11}$$

Proof: See Appendix A. □

Note that the prior distribution of δ in $\widetilde{M}_K$, appearing on the right-hand side of (10), is actually equal to $\pi_{\underline{\delta}}(\delta|M_K)$, for the transformation g does not affect the parameters collected in δ.

We end this section by formulating our basic result in Lemma 3, with Corollary 4 following immediately.

Lemma 3. *For a given $j\in\{1,2,\ldots,n\}$, the prior distribution of $\lambda_j^{(K)}$ under M_K, and the one of the corresponding parameter $\lambda_{1,j}$ under M_1 are compatible iff*

$$\pi_{\underline{\lambda_{1,j}}}\left(\lambda_{1,j}|M_1\right)\propto\prod_{i=1}^{K}\pi_{\underline{\lambda_{i,j}}}\left(\lambda_{1,j}|M_K\right). \tag{12}$$

Proof:
Employing (7) and (11), we proceed as follows:

$$\begin{aligned}\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1) &= \pi_{\underline{\lambda_{1,j}}|\underline{\tau_j}}\left(\lambda_{1,j}|\tau_j = 0_{[(K-1)\times 1]}, \widetilde{M}_K\right) \\ &\propto \pi_{\underline{\lambda_{1,j}},\underline{\tau_j}}\left(\lambda_{1,j}, \tau_j = 0_{[(K-1)\times 1]}|\widetilde{M}_K\right) \\ &\propto \pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_K)\left[\prod_{i=2}^{K}\pi_{\underline{\lambda_{i,j}}}(\lambda_{1,j}|M_K)\right] = \\ &= \prod_{i=1}^{K}\pi_{\underline{\lambda_{i,j}}}(\lambda_{1,j}|M_K).\end{aligned}$$

□

Corollary 4. *For a given $j \in \{1, 2, \ldots, n\}$, if all densities $\pi_{\underline{\lambda_{i,j}}}(\cdot|M_K)$, $i = 1, 2, \ldots, K$, are the same, i.e., $\pi_{\underline{\lambda_{i,j}}}(x|M_K) = \pi_{\underline{\lambda_{1,j}}}(x|M_K)$, $x \in \mathbb{R}$, then (12) reduces to*

$$\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1) \propto \left[\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_K)\right]^K. \tag{13}$$

3 Specific results

The relations presented in (12) and (13) may prompt one, quite instinctively, to consider some exponential family for specyfing the prior densities of $\lambda_{i,j}$'s under M_K, since such an approach would yield the same type of the prior distribution for $\lambda_{1,j}$ under M_1. What remains then is to determine the relationships between the hyperparameters of all the relevant densities (belonging to a given exponential family).
In the subsections below we focus our attention on three exponential families: the normal, inverse gamma, and gamma distributions, which, for their property of (conditional) conjugacy, are commonly entertained in Bayesian statistical modeling. In each case, we apply Lemma 2 and Corollary 4 to derive explicit formulae relating the hyperparameters of the general and the nested model. Throughout the section we fix the index $j \in \{1, 2, \ldots, n\}$, and, for the sake of transparency, drop it from the notation (e.g., writing λ_i instead of $\lambda_{i,j}$).

3.1 Normal priors

The following proposition establishes the compatibility conditions upon the normality of λ_i's in the mixture model.

Proposition 5. *Suppose that each λ_i ($i = 1, 2, \dots, K$) under M_K follows a univariate normal distribution with mean $m_i^{(K)}$ and variance $v_i^{(K)}$:*

$$\pi_{\underline{\lambda_i}}(\lambda_i|M_K) = f_N\left(\lambda_i|m_i^{(K)}, v_i^{(K)}\right).$$

Then, the compatible prior for λ_1 under M_1 is the normal distribution with mean $m^{(1)}$ and variance $v^{(1)}$:

$$\pi_{\underline{\lambda_1}}(\lambda_1|M_1) = f_N\left(\lambda_1|m^{(1)}, v^{(1)}\right),$$

where

$$m^{(1)} = \frac{\sum_{i=1}^{K} \frac{m_i^{(K)}}{v_i^{(K)}}}{\sum_{i=1}^{K} \frac{1}{v_i^{(K)}}} \tag{14}$$

and

$$v^{(1)} = \left(\sum_{i=1}^{K} \frac{1}{v_i^{(K)}}\right)^{-1}. \tag{15}$$

Alternatively, under precision-parametrized normal densities, if

$$\pi_{\underline{\lambda_i}}(\lambda_i|M_K) = f_N\left(\lambda_i|m_i^{(K)}, (\breve{v}_i^{(K)})^{-1}\right), \qquad \breve{v}_i^{(K)} \equiv (v_i^{(K)})^{-1},$$

for each $i = 1, 2, \dots, K$, then

$$\pi_{\underline{\lambda_1}}(\lambda_1|M_1) = f_N\left(\lambda_1|m^{(1)}, (\breve{v}^{(1)})^{-1}\right),$$

where

$$m^{(1)} = \frac{\sum_{i=1}^{K} \breve{v}_i^{(K)} m_i^{(K)}}{\sum_{i=1}^{K} \breve{v}_i^{(K)}} \tag{16}$$

and

$$\breve{v}^{(1)} = \sum_{i=1}^{K} \breve{v}_i^{(K)}. \tag{17}$$

Proof. See Appendix B. □

Following immediately from Proposition 5, the corollary below provides expressions for $m^{(1)}$, $v^{(1)}$ and $\breve{v}^{(1)}$ upon the component-wise equality of hyperparameters under M_K.

Corollary 6. *i) If* $m_1^{(K)} = m_2^{(K)} = \ldots = m_K^{(K)} \equiv m^{(K)}$*, then*

$$m^{(1)} = m^{(K)}. \tag{18}$$

ii) If $v_1^{(K)} = v_2^{(K)} = \ldots = v_K^{(K)} \equiv v^{(K)}$ *or, equivalently,* $\breve{v}_1^{(K)} = \ldots = \breve{v}_K^{(K)} \equiv \breve{v}^{(K)}$*, then*

$$m^{(1)} = \frac{1}{K}\sum_{i=1}^{K} m_i^{(K)}, \tag{19}$$

$$v^{(1)} = \frac{1}{K} v^{(K)} \tag{20}$$

and

$$\breve{v}^{(1)} = K\breve{v}^{(K)}. \tag{21}$$

According to (14) and (16), the mean of the compatible (normal) prior of λ_1 under the nested model constitutes a weighted sum of the corresponding means in the mixture model:

$$m^{(1)} = \sum_{i=1}^{K} w_i m_i^{(K)},$$

with the weights given by

$$w_i = \frac{(v_i^{(K)})^{-1}}{\sum_{k=1}^{K} (v_k^{(K)})^{-1}} = \frac{\breve{v}_i^{(K)}}{\sum_{k=1}^{K} \breve{v}_k^{(K)}}, \qquad i = 1, 2, \ldots K.$$

The result collapses either to a simple average of the means (under equal variances $v_i^{(K)}$; see (19)), or, eventually, to the very mean $m^{(K)}$, should the means coincide in all the priors $\pi_{\underline{\lambda_i}}(\lambda_i|M_K)$, $i = 1, 2, \ldots, K$; see (18).

As regards the relationship between the dispersion of the priors, from (15) it follows that the variance $v^{(1)}$ in the compatible prior of λ_1 under M_1 amounts to a K-th of the harmonic mean of the individual variances $v_i^{(K)}$, $i = 1, 2, \ldots, K$. The result immediately translates to the relation between the corresponding precisions, in terms of which $\breve{v}^{(1)}$ in the reduced model should be the sum of the precisions specified in the general construction; see (17). Under the special case of equal prior variances of all λ_i's in M_K, the resulting variance of λ_1 in M_1 reduces to a K-th of the one assumed within the mixture model; see (20). Equivalently, the precision $\breve{v}^{(1)}$ is K times the one predetermined for λ_i's, thereby growing proportionally to the number of the mixture components; see (21).

We end this subsection by noticing that under the assumptions of Proposition

5 it is only possible to determine the hyperparameters in the single-component specification, based on the ones prespecified in the mixture model, and not the reverse. However, adopting an additional assumption of the equal prior means and, simultaneously, variances of λ_i's under M_K, allows one to predetermine the hyperparameters for λ_1 in the nested model first (i.e., $m^{(1)}$ and $v^{(1)}$), and then the ones in the general specification (i.e., $m^{(K)}$ and $v^{(K)}$), employing (18) and (20) (or, equivalently, (21)). The latter idea appears to gain particular importance while considering models with various number of the mixture components: M_K with $K \in \{K_{min}, K_{min}+1, \ldots, K_{max}\} = \mathbb{K}$, $K_{min} \geq 2$, along the single-component structure, M_1. Since the latter constitutes a special case of all the mixture models under consideration, one may naturally be prompted to set the hyperparameters under M_1 first, and then invoke (18) and (20) (or, (21)) to calculate compatible values of $m^{(K)}$ and $v^{(K)}$ (or, $\breve{v}^{(K)}$) for each $K \in \mathbb{K}$. Intuitively, though not in the sense of Definition 1, such an approach would endow the priors of all the models with some sort of compatibility, by means of ensuring prior compatibility of the single-component model with each of the mixture specifications individually.

3.2 Inverse gamma priors

We move on to deriving the compatibility conditions under setting inverse gamma priors for λ_i's in the mixture model.

Proposition 7. *Suppose that each λ_i ($i = 1, 2, \ldots, K$) under M_K follows an inverse gamma distribution with shape parameter $a_i^{(K)} > 0$ and scale parameter $b_i^{(K)} > 0$:*

$$\pi_{\underline{\lambda_i}}(\lambda_i|M_K) = f_{IG}\left(\lambda_i|a_i^{(K)}, b_i^{(K)}\right)$$
$$= \frac{1}{\left(b_i^{(K)}\right)^{a_i^{(K)}} \Gamma\left(a_i^{(K)}\right)} (\lambda_i)^{-\left(a_i^{(K)}+1\right)} \exp\left\{-\frac{1}{b_i^{(K)}\lambda_i}\right\}.$$

Then, the compatible prior for λ_1 under M_1 is the inverse gamma distribution with shape parameter $a^{(1)}$ and scale parameter $b^{(1)}$:

$$\pi_{\underline{\lambda_1}}(\lambda_1|M_1) = f_{IG}\left(\lambda_1|a^{(1)}, b^{(1)}\right),$$

where

$$a^{(1)} = \sum_{i=1}^{K} a_i^{(K)} + K - 1 \tag{22}$$

and

$$b^{(1)} = \left(\sum_{i=1}^{K} \frac{1}{b_i^{(K)}}\right)^{-1}. \tag{23}$$

Proof. See Appendix C. □

The formulae for $a^{(1)}$ and $b^{(1)}$ in the special cases of component-wise equal hyperparameters under the general model follow directly from Proposition 7 and are stated in the corollary below.

Corollary 8. *i) If* $a_1^{(K)} = a_2^{(K)} = \cdots = a_K^{(K)} \equiv a^{(K)}$*, then*

$$a^{(1)} = Ka^{(K)} + K - 1. \tag{24}$$

ii) If $b_1^{(K)} = b_2^{(K)} = \ldots = b_K^{(K)} \equiv b^{(K)}$*, then*

$$b^{(1)} = \frac{1}{K} b^{(K)}. \tag{25}$$

The relationship between the shape parameters, given by (22), suggests that $a^{(1)}$ is an increasing function of the number of the mixture components (partly on account of its formula involving the sum of $a_i^{(K)}$'s), whereas the scale parameters, $b_i^{(K)}$'s and $b^{(1)}$, are interrelated in the same fashion as the variances in the case of the normal priors, examined in the previous subsection (see (23) and (15)).
Similarly to the previous one, Proposition 7 enables one to derive compatible values of the hyperparameters under M_1, based on the ones prespecified under the mixture model, unless these are held equal across the mixture components (see Corollary i). Turning to the special case of $a_1^{(K)} = a_2^{(K)} = \ldots = a_K^{(K)} \equiv a^{(K)}$, let us transform (24) into

$$a^{(K)} = \frac{a^{(1)} - K + 1}{K}, \tag{26}$$

which would be of use once we were to establish the compatible prior in M_K, based on the predetermined value of the relevant hyperparameter in M_1. Interestingly, to guarantee the positivity of $a^{(K)}$ (as a shape parameter of an inverse gamma distribution) it requires that

$$a^{(1)} > K - 1, \tag{27}$$

which explicitly takes the number of mixture components into account. Now, evoke the context of handling models M_K with various $K \in \mathbb{K}$, as outlined at the end of the previous subsection. In order to ascertain the prior under each of them coherently with the one prespecified for the single-component model, the condition

$$a^{(1)} > K_{max} - 1 \tag{28}$$

must be satisfied. As long as (28) holds, the hyperparameters $a^{(K)}$ calculated through (26) are positive for all $K \in \mathbb{K}$. Taking these remarks into account, it emerges that once models with a different number of the components are under consideration, it is crucial to fix *a priori* its maximum, K_{max}. With that provided, one proceeds to setting $a^{(1)}$ in compliance with (28), and then to determining $a^{(K)}$ via (26) for each $K \in \mathbb{K}$. Incidentally, note that the issue pertains only to the shape parameters, while reconciling the scale parameters: $b^{(1)}$ and $b^{(K)}$ (under $b_1^{(K)} = b_2^{(K)} = \ldots = b_K^{(K)} \equiv$ $\equiv b^{(K)}$) for each $K \in \mathbb{K}$, does not give rise to similar concerns.

3.3 Gamma priors

Generally speaking, in some applications it is preferred to employ the gamma distribution (rather than its inverse alternative) to specify the prior. Therefore, we devote the present subsection to provide the compatibility conditions also in the case gamma priors are assumed for all λ_i's in the mixture model.

Proposition 9. *Suppose that each λ_i $(i = 1, 2, \ldots, K)$ under M_K follows a gamma distribution with shape parameter $\breve{a}_i^{(K)} > 0$ and scale parameter $\breve{b}_i^{(K)} > 0$:*

$$\pi_{\underline{\lambda_i}}(\lambda_i|M_K) = f_G\left(\lambda_i|\breve{a}_i^{(K)}, \breve{b}_i^{(K)}\right)$$
$$= \frac{\left(\breve{b}_i^{(K)}\right)^{\breve{a}_i^{(K)}}}{\Gamma\left(\breve{a}_i^{(K)}\right)}(\lambda_i)^{\breve{a}_i^{(K)}-1} \exp\left\{-\breve{b}_i^{(K)}\lambda_i\right\}.$$

Then, the compatible prior for λ_1 under M_1 is the gamma distribution with shape parameter $\breve{a}^{(1)}$ and scale parameter $\breve{b}^{(1)}$:

$$\pi_{\underline{\lambda_1}}(\lambda_1|M_1) = f_G\left(\lambda_1|\breve{a}^{(1)}, \breve{b}^{(1)}\right),$$

where

$$\breve{a}^{(1)} = \sum_{i=1}^{K} \breve{a}_i^{(K)} - K + 1 \tag{29}$$

and

$$\breve{b}^{(1)} = \sum_{i=1}^{K} \breve{b}_i^{(K)}. \tag{30}$$

Proof. See Appendix D. □

Similarly as in the previous subsections, and following directly from Proposition 9, the corollary below delivers expressions for $\breve{a}^{(1)}$ and $\breve{b}^{(1)}$ under the special cases of component-wise equal hyperparameters in the mixture model.

Corollary 10. *i) If $\breve{a}_1^{(K)} = \breve{a}_2^{(K)} = \ldots = \breve{a}_K^{(K)} \equiv \breve{a}^{(K)}$, then*

$$\breve{a}^{(1)} = K\breve{a}^{(K)} - K + 1. \tag{31}$$

ii) If $\breve{b}_1^{(K)} = \breve{b}_2^{(K)} = \ldots = \breve{b}_K^{(K)} \equiv \breve{b}^{(K)}$, *then*

$$\breve{b}^{(1)} = K\breve{b}^{(K)}. \tag{32}$$

With regard to the relationship between the shape parameters, in general, (29) reveals no evident monotonic dependency of $\breve{a}^{(1)}$ upon the number of mixture components.

In the special case of the component-wise equal $\breve{a}_i^{(K)}$'s, it is easily gathered from (31) that $\breve{a}^{(1)} = K(\breve{a}^{(K)}-1)+1$, which implies $\breve{a}^{(1)}$ may be constant in K (under $\breve{a}^{(K)} = 1$), or increasing ($\breve{a}^{(K)} > 1$), or decreasing ($\breve{a}^{(K)} < 1)K$.

As far as the scale parameters are concerned, they follow the pattern of the precisions entertained under the precision-parametrized normal priors in Subsection 3.1 (see (30) and (17)), rather than the variances, which was the case under the inverse gamma priors. Contrary to the inverse gamma priors analyzed previously, working under the gamma distributions provides an easy route to establishing compatible values of the shape parameters once, again, models M_K with various $K \in \mathbb{K}$ are at hand, and, given the number of components, all the hyperparameters $\breve{a}_i^{(K)}$'s are held equal. To this end, transform (31) and (32), respectively, into

$$\breve{a}^{(K)} = \frac{\breve{a}^{(1)} + K - 1}{K} \tag{33}$$

and

$$\breve{b}^{(K)} = \frac{1}{K}\breve{b}^{(1)}. \tag{34}$$

Setting any $\breve{a}^{(1)} > 0$ in (33) yields a positive value of $\breve{a}^{(K)}$ for any $K \in \{2, 3, \ldots\}$. Hence, the approach disposes of *a priori* fixing the maximum number of components, otherwise necessitated under the inverse gamma framework.

4 Compatibility of constrained prior distributions

In the foregoing, only unconstrained priors, given by (2) and (3), under all the models have been considered. However, in practice, it may be that some restrictions are to be imposed on the parameters of the mixture model, usually aiming at ensuring the identifiability of the mixture components or, possibly in addition to that, some sort of regularity, such as the second-order stationarity (in the time series framework). Therefore, in the present section, we study the way in which introducing such parametric constraints into the mixture model's prior affects the general results stated in Lemma 3 and Corollary 4.

4.1 Priors with identifiability constraints

There has already been a large variety of techniques advanced in the literature to exert identifiability of the mixture model's components, each procedure designed to tackle the widely-recognized label switching issue, an inherent ailment of the mixture modeling. For a review and more recent studies in the field we refer the reader to, e.g., Jasra et al. (2005), Marin et al. (2005), Frühwirth- Schnatter (2006), Yao (2012a), Yao (2012b), and the references therein. The most straightforward method (though not universally recommended, according to the cited authors) consists in imposing an inequality constraint upon the coordinates of the vector $\lambda_j^{(K)}$ for a given $j \in \{1, 2, \ldots, n\}$, such as

$$\lambda_{1,j} \leq \lambda_{2,j} \leq \ldots \leq \lambda_{K,j}. \tag{35}$$

(Notice that the subscript j is henceforth reintroduced in the notation). We stress that it is the strict-inequalities variant of (35) that is usually engaged in the literature, thereby actually prohibiting the single-component structure from nesting itself within the mixture model. Admittedly, such an approach is entirely valid within the subjective setting, which, obviously, does not necessitate establishing any relation between the models under consideration, their priors included, even if such a one is conceivable. However, aiming at ensuring the prior compatibility between the single-component model and the mixture model, with the latter's prior constrained, does require allowing for the weak inequalities in (35), for otherwise the former could not be obtained from the latter via conditioning upon $\lambda_{1,j} = \lambda_{2,j} = \ldots = \lambda_{K,j}$, $j = 1, 2, \ldots, n$. Finally, note that the distinction between the weak and the strict inequalities within a continuous random variables framework is hardly a matter of concern.

With no loss of generality we assume that the identifiability restriction is imposed on the prior of $\lambda_1^{(K)}$, i.e., for $j = 1$, whereas the priors of the remaining $\lambda_j^{(K)}$'s ($j = 2, 3, \ldots, n$) are unconstrained and coincide with (4). The prior for $\lambda_1^{(K)}$ can be written as

$$\pi_{\underline{\lambda_1^{(K)}}}\left(\lambda_1^{(K)}|M_K\right) \propto \left[\prod_{i=1}^{K} \pi_{\underline{\lambda_{i,1}}}\left(\lambda_{i,1}|M_K\right)\right] \mathbb{I}_{C_K}\left(\lambda_1^{(K)}\right), \tag{36}$$

where

$$C_K = \{(c_1\ c_2\ \ldots\ c_K)' \in \mathbb{R}^K : c_1 \leq c_2 \leq \ldots \leq c_K\}$$

and $\mathbb{I}_{C_K}(\cdot)$ represents the indicator function of the set C_K. Incidentally, note a slight abuse of notation in (36), for $\pi_{\underline{\lambda_{i,1}}}(\lambda_{i,1}|M_K)$ is actually no longer the marginal prior of $\lambda_{i,1}$, which is due to the inequality constraint introducing stochastic dependency between the coordinates of $\lambda_1^{(K)}$.

Proceeding along the same lines of reasoning as presented in Section 2, we rewrite

(36) under the reparametrized model, $\widetilde{M}_K$:

$$\pi_{\underline{\widetilde{\lambda}_1^{(K)}}}\left(\widetilde{\lambda}_1^{(K)}|\widetilde{M}_K\right) = \pi_{\underline{\lambda_1^{(K)}}}\left(g^{-1}\left(\widetilde{\lambda}_1^{(K)}\right)|M_K\right)$$
$$\propto \pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_K)\left[\prod_{i=2}^{K}\pi_{\underline{\lambda_{i,1}}}(\tau_{i,1}+\lambda_{1,1}|M_K)\right]\mathbb{I}_{C^+_{K-1}}(\tau_1), \qquad (37)$$

where

$$C^+_{K-1} = \{(c_1\ c_2\ \ldots\ c_{K-1})' \in \mathbb{R}^{K-1} : 0 \le c_1 \le c_2 \le \ldots \le c_{K-1}\}$$

and $\tau_1 = (\tau_{2,1}\ \tau_{3,1}\ \ldots\ \tau_{K,1})'$, $\tau_{i,1} = \lambda_{i,1} - \lambda_{1,1}$ $(i = 2, 3, ..., K)$, so that the presence of $\mathbb{I}_{C^+_{K-1}}(\tau_1)$ in (37) is equivalent to restricting the contrasts with the inequality $0 \le \tau_{2,1} \le \tau_{3,1} \le \ldots \le \tau_{K,1}$. Now, recognizing that $\mathbb{I}_{C^+_{K-1}}(0_{[(K-1)\times 1]}) = 1$, and following the proof of Lemma 3 we obtain:

$$\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1) = \pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1 = 0_{[(K-1)\times 1]}, \widetilde{M}_K)$$
$$\propto \pi_{\underline{\lambda_{1,1}},\underline{\tau_1}}(\lambda_{1,j}, \tau_1 = 0_{[(K-1)\times 1]}|\widetilde{M}_K)$$
$$\propto \pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_K)\left[\prod_{i=2}^{K}\pi_{\underline{\lambda_{i,1}}}(\lambda_{1,1}|M_K)\right]\mathbb{I}_{C^+_{K-1}}(0_{[(K-1)\times 1]}) =$$
$$= \prod_{i=1}^{K}\pi_{\underline{\lambda_{i,1}}}(\lambda_{1,1}|M_K),$$

which coincides with the result displayed in the lemma. Hence, we conclude that constraining the mixture model's prior with an identifiability restriction does not affect the compatibility conditions stated in Lemma 3 and Corollary 4.

4.2 Priors with regularity constraints

Another common type of parametric restrictions introduced into statistical models are the ones enforcing some sort of regularity, arising from the theory underlying the phenomenon at hand or being of a rather technical nature (e.g., ensuring the second-order stationarity in the time series framework). Therefore, we move on to establishing the way in which a regularity restriction imposed upon the mixture model's prior translates into the form of the compatible prior under the single-component specification.

Let $\zeta_K(\cdot) : \Theta^{(K)} \to \mathbb{R}$ be such a function of $\theta^{(K)}$ that the regularity constraint under M_K is satisfied if and only if $\zeta_K(\theta^{(K)}) \in R_K \subset \mathbb{R}$ (or, equivalently, $\mathbb{I}_{R_K}\left\{\zeta_K(\theta^{(K)})\right\} = 1$), and $\zeta_K(\theta^{(K)})$ becomes invariant with respect to η under the reducing restrictions given by (8).

Rewriting $\theta^{(K)} = \left(\delta'\ \lambda_1^{(K)\prime}\ \lambda_2^{(K)\prime}\ \ldots\ \lambda_n^{(K)\prime}\ \eta'\right)'$ as $\theta^{(K)} = (\delta'\ \ \lambda^{(K)\prime}\ \ \eta')'$ with

$\lambda^{(K)} = \left(\lambda_1^{(K)\prime}\ \lambda_2^{(K)\prime}\ \ldots\ \lambda_n^{(K)\prime}\right)^\prime$, and assuming prior independence (though only up to the regularity restriction), the constrained prior under the mixture model presents itself as

$$\pi_{\underline{\theta^{(K)}}}(\theta^{(K)}|M_K) \propto \pi_{\underline{\delta}}(\delta|M_K)\pi_{\underline{\lambda^{(K)}}}(\lambda^{(K)}|M_K)\pi_{\underline{\eta}}(\eta|M_K)\mathbb{I}_{R_K}\left\{\zeta_K(\theta^{(K)})\right\} =$$
$$= \pi_{\underline{\delta}}(\delta|M_K)\left[\prod_{j=1}^{n}\pi_{\underline{\lambda_j^{(K)}}}(\lambda_j^{(K)}|M_K)\right]\pi_{\underline{\eta}}(\eta|M_K)\mathbb{I}_{R_K}\left\{\zeta_K(\theta^{(K)})\right\},$$

with each $\pi_{\underline{\lambda_j^{(K)}}}(\lambda_j^{(K)}|M_K)$ $(j = 1, 2, \ldots, n)$ being given by (4). As regards deriving from the above expression the compatible prior under the single-component model, one conjectures that it is also to be constrained with some restriction, say $\zeta_1(\theta^{(1)}) \in R_1 \subset \mathbb{R}$. Although a precise relation between $\zeta_1(\cdot) : \Theta^{(1)} \to \mathbb{R}$ and $\zeta_K(\cdot)$ is yet to be specified, we shall write a prototypical form, so to say, of the prior under M_1:

$$\pi_{\underline{\theta^{(1)}}}(\theta^{(1)}|M_1) \propto \pi_{\underline{\delta}}(\delta|M_1)\left[\prod_{j=1}^{n}\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1)\right]\mathbb{I}_{R_1}\left\{\zeta_1(\theta^{(1)})\right\}. \tag{38}$$

Further, let us recast M_K into $\widetilde{M}_K$ (with the transform g affecting only $\lambda_j^{(K)}$'s, as in Section 2), so that

$$\pi_{\underline{\widetilde{\theta}^{(K)}}}(\widetilde{\theta}^{(K)}|\widetilde{M}_K) \propto \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)\pi_{\underline{\widetilde{\lambda}^{(K)}}}(\widetilde{\lambda}^{(K)}|\widetilde{M}_K)\pi_{\underline{\eta}}(\eta|\widetilde{M}_K)$$
$$\times\, \mathbb{I}_{R_K}\left\{\zeta_K\left(\delta, g^{-1}(\widetilde{\lambda}_1^{(K)}), g^{-1}(\widetilde{\lambda}_2^{(K)}), \ldots, g^{-1}(\widetilde{\lambda}_n^{(K)}), \eta\right)\right\},$$

where

$$\widetilde{\lambda}^{(K)} = (\widetilde{\lambda}_1^{(K)\prime}\ \widetilde{\lambda}_2^{(K)\prime}\ \ldots\ \widetilde{\lambda}_n^{(K)\prime})^\prime,$$
$$\pi_{\underline{\delta}}(\delta|\widetilde{M}_K) = \pi_{\underline{\delta}}(\delta|M_K),$$
$$\pi_{\underline{\eta}}(\eta|\widetilde{M}_K) = \pi_{\underline{\eta}}(\eta|M_K)$$

and

$$\pi_{\underline{\widetilde{\lambda}^{(K)}}}(\widetilde{\lambda}^{(K)}|\widetilde{M}_K) = \prod_{j=1}^{n}\pi_{\underline{\widetilde{\lambda}_j^{(K)}}}(\widetilde{\lambda}_j^{(K)}|\widetilde{M}_K).$$

Employing the end result of (7) into $\pi_{\underline{\widetilde{\theta}^{(K)}}}(\widetilde{\theta}^{(K)}|\widetilde{M}_K)$, one obtains

$$\pi_{\underline{\widetilde{\theta}^{(K)}}}(\widetilde{\theta}^{(K)}|\widetilde{M}_K) \propto \pi_{\underline{\delta}}(\delta|M_K)\left[\prod_{j=1}^{n}\left(\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_K)\prod_{i=2}^{K}\pi_{\underline{\lambda_{i,j}}}(\tau_{i,j}+\lambda_{1,j}|M_K)\right)\right]$$
$$\times \pi_{\underline{\eta}}(\eta|M_K)\mathbb{I}_{R_K}\left\{\zeta_K\left(\delta, g^{-1}(\widetilde{\lambda}_1^{(K)}), g^{-1}(\widetilde{\lambda}_2^{(K)}),\ldots,g^{-1}(\widetilde{\lambda}_n^{(K)}),\eta\right)\right\}.$$

Now, notice that under $\tau_j = 0_{[(K-1)\times 1]}$, we get $g^{-1}(\widetilde{\lambda}_j^{(K)}) = \lambda_{1,j}\iota_K$, with $\iota_K = (1\ 1\ \ldots\ 1)' \in \mathbb{R}^K$ and $j = 1, 2, \ldots, n$. Finally, the compatible prior distribution under M_1 is derived:

$$\begin{aligned}\pi_{\underline{\theta^{(1)}}}(\theta^{(1)}|M_1) \propto\ & \pi_{\underline{\delta}}(\delta|M_K)\\ & \times \pi_{\underline{\widetilde{\lambda}^{(K)}}|\underline{\tau_2},\underline{\tau_3},\ldots,\underline{\tau_n}}(\widetilde{\lambda}^{(K)}|\tau_2=\tau_3=\ldots=\tau_n=0_{[(K-1)\times 1]},\widetilde{M}_K)\\ & \times \mathbb{I}_{R_K}\left\{\zeta_K\left(\delta,\lambda_{1,1}\iota_K,\lambda_{1,2}\iota_K,\ldots,\lambda_{1,n}\iota_K,\eta\right)\right\} =\\ & = \pi_{\underline{\delta}}(\delta|M_K)\left[\prod_{j=1}^{n}\prod_{i=1}^{K}\pi_{\underline{\lambda_{i,j}}}(\lambda_{1,j}|M_K)\right]\\ & \times \mathbb{I}_{R_K}\left\{\zeta_K\left(\delta,\lambda_{1,1}\iota_K,\lambda_{1,2}\iota_K,\ldots,\lambda_{1,n}\iota_K,\eta\right)\right\}.\end{aligned}$$

To reconcile the above expression with (38), the following conditions must hold simultaneously:

$$\pi_{\underline{\delta}}(\delta|M_1) = \pi_{\underline{\delta}}(\delta|M_K), \tag{39}$$

$$\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1) \propto \prod_{i=1}^{K}\pi_{\underline{\lambda_{i,j}}}(\lambda_{1,j}|M_K), \tag{40}$$

$$\mathbb{I}_{R_1}\left\{\zeta_1(\theta^{(1)})\right\} = 1 \Leftrightarrow \mathbb{I}_{R_K}\left\{\zeta_K\left(\delta,\lambda_{1,1}\iota_K,\lambda_{1,2}\iota_K,\ldots,\lambda_{1,n}\iota_K,\eta\right)\right\} = 1. \tag{41}$$

Note that (39) and (40) coincide with (5) and (12), respectively. Hence, from (39)-(41) it follows that in order to design a compatible prior under the single-component model one needs to:

1. Comply with the rules formulated for the case of unconstrained priors; see (5) and Lemma 3.

2. Restrain the single-component model's prior with a restriction equivalent to the one restraining the mixture model's prior under the nesting restrictions, given by (8).

5 Example: A compatible prior structure for a class of stationary Markov-switching AR(2) models

Consider the following K-state Markov-switching AR(2) model:

$$y_t = \alpha_{S_t} + \phi_{S_t,1} y_{t-1} + \phi_{S_t,2} y_{t-2} + \sigma_{S_t} \varepsilon_t, \tag{42}$$

where $\varepsilon_t \sim iiN(0,1)$ and the sequence $\{S_t\}$ forms a homogeneous and ergodic Markov chain with finite state-space $\mathbb{S} = \{1, 2, \ldots, K\}$ and transition probabilities $\eta_{ij} \equiv \Pr(S_t = j | S_{t-1} = i)$, arrayed in transition matrix $P = [\eta_{ij}]_{i,j=1,2,\ldots,K}$. Adopting the convention introduced by Krolzig (1997), we refer to (42) as the MSIAH(K)-AR(2) model (or, M_K, in short), which indicates allowing all the parameters to feature Markovian breaks, i.e., the intercept, the autoregressive coefficients and the error term's variance. Let $\alpha^{(K)} = (\alpha_1 \ \alpha_2 \ \ldots \ \alpha_K)'$, $\phi_1^{(K)} = (\phi_{1,1} \ \phi_{2,1} \ \ldots \ \phi_{K,1})'$, $\phi_2^{(K)} = (\phi_{1,2} \ \phi_{2,2} \ \ldots \ \phi_{K,2})'$, $\varsigma^{(K)} = (\sigma_1^{-2} \ \sigma_2^{-2} \ \ldots \ \sigma_K^{-2})'$, and η be structured as described in Section 2, so that

$$\theta^{(K)} = \left(\alpha^{(K)\prime} \ \phi_1^{(K)\prime} \ \phi_2^{(K)\prime} \ \varsigma^{(K)\prime} \ \eta' \right)'.$$

The model under consideration generalizes the following AR(2) specification (hereafter denoted by M_1):

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \sigma \varepsilon_t \tag{43}$$

in that M_K introduces discrete changes into each of the four parameters of M_1 (grouped in $\theta^{(1)} = (\alpha \ \phi_1 \ \phi_2 \ \sigma^{-2})'$). Based on the results provided by Francq and Zakoïan (2001), for the MSIAH(K)-AR(2) process to be nonanticipative (i.e., causal) and second-order stationary it suffices that

$$\rho(P_2) < 1, \tag{44}$$

where $\rho(P_2)$ signifies the spectral radius of matrix P_2 defined as

$$P_2 = \begin{pmatrix} \eta_{11}(\Phi_1 \otimes \Phi_1) & \eta_{21}(\Phi_1 \otimes \Phi_1) & \cdots & \eta_{K1}(\Phi_1 \otimes \Phi_1) \\ \eta_{12}(\Phi_2 \otimes \Phi_2) & \eta_{22}(\Phi_2 \otimes \Phi_2) & \cdots & \eta_{K2}(\Phi_2 \otimes \Phi_2) \\ \vdots & \vdots & & \vdots \\ \eta_{1K}(\Phi_K \otimes \Phi_K) & \eta_{2K}(\Phi_K \otimes \Phi_K) & \cdots & \eta_{KK}(\Phi_K \otimes \Phi_K) \end{pmatrix}, \tag{45}$$

with

$$\Phi_k = \begin{pmatrix} \phi_{k,1} & \phi_{k,2} \\ 1 & 0 \end{pmatrix}, \qquad k = 1, 2, \ldots, K,$$

and $\otimes$ denoting the matrix tensor product. Assuming the mutual independence of $\theta^{(K)}$'s individual components, the prior under M_K can be written as

$$\pi(\theta^{(K)}|M_K) = \pi(\alpha^{(K)}|M_K)\pi(\phi_1^{(K)}|M_K)\pi(\phi_2^{(K)}|M_K) \\ \times \pi(\varsigma^{(K)}|M_K)\pi(\eta|M_K)\mathbb{I}_{R_K}\{\rho(P_2)\},$$

where $R_K = [0, 1)$. Note that we simplified the notation by dropping the subscripts indexing densities, and write, generally, $\pi(\omega)$ instead of $\pi_{\underline{\omega}}(\omega)$. Similarly, the prior under M_1 is given by

$$\pi(\theta^{(1)}|M_1) = \pi(\alpha|M_1)\pi(\phi_1|M_1)\pi(\phi_2|M_1) \\ \times \pi(\sigma^{-2}|M_1)\mathbb{I}_{R_1}\{\zeta_1(\theta^{(1)})\}.$$

Notice that, for the sake of exposition, we do not impose any identifiability restriction upon $\pi(\theta^{(K)}|M_K)$, though we stress that it would not alter the following considerations (see Subsection 4.1).
To derive the specific forms of $\zeta_1(\theta^{(1)})$ and R_1, complying with the compatibility condition given by (41), one needs to ponder (44) under the equality restrictions: $\phi_{1,1} = \phi_{2,1} = \ldots = \phi_{K,1} \equiv \phi_1$ and $\phi_{1,2} = \phi_{2,2} = \ldots = \phi_{K,2} \equiv \phi_2$. (Notice that the switching intercepts, $\alpha^{(K)}$, and the error term's precisions, $\varsigma^{(K)}$, do not need to be restricted with the nesting equalities, in the process). With that provided, the matrices Φ_k's collapse into

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix},$$

which coincides with the companion matrix for the AR(2) process defined in (43). Supplanting Φ_k's with Φ in (45) we obtain

$$P_2 = \begin{pmatrix} \eta_{11}(\Phi\otimes\Phi) & \eta_{21}(\Phi\otimes\Phi) & \cdots & \eta_{K1}(\Phi\otimes\Phi) \\ \eta_{12}(\Phi\otimes\Phi) & \eta_{22}(\Phi\otimes\Phi) & \cdots & \eta_{K2}(\Phi\otimes\Phi) \\ \vdots & \vdots & & \vdots \\ \eta_{1K}(\Phi\otimes\Phi) & \eta_{2K}(\Phi\otimes\Phi) & \cdots & \eta_{KK}(\Phi\otimes\Phi) \end{pmatrix} = P'\otimes\Phi\otimes\Phi.$$

Then $\rho(P_2) = \rho(P'\otimes\Phi\otimes\Phi) = \rho(P')[\rho(\Phi)]^2 = [\rho(\Phi)]^2$, for P is a stochastic matrix. Finally,

$$\mathbb{I}_{R_K}\{\rho(P_2)\} = 1 \Leftrightarrow \mathbb{I}_{R_K}\left\{[\rho(\Phi)]^2\right\} = 1 \\ \Leftrightarrow \mathbb{I}_{R_K}\{\rho(\Phi)\} = 1.$$

The latter expression requires that the maximum absolute eigenvalue of Φ be less than one, which is equivalent to the well-known condition for the AR(2) process to be nonanticipative and second-order stationary, necessitating all eigenvalues of the

companion matrix to fall within the interval $(-1, 1)$. Therefore, we assume that $\zeta_1(\theta^{(1)}) := \rho(\Phi)$ and $R_1 = R_K = [0, 1)$.
As regards particular choice for the individual densities comprising $\pi(\theta^{(K)}|M_K)$, while keeping to the assumptions stated in Section 2, we follow a typical framework by setting

normal distributions for the coordinates of $\alpha^{(K)}$, $\phi_1^{(K)}$ and $\phi_2^{(K)}$:

$$\pi(\alpha^{(K)}|M_K) = \prod_{i=1}^{K} f_N\left(\alpha_i|m_\alpha^{(K)}, (\breve{v}_\alpha^{(K)})^{-1}\right), \tag{46}$$

$$\pi(\phi_1^{(K)}|M_K) = \prod_{i=1}^{K} f_N\left(\phi_{i,1}|m_{\phi_1}^{(K)}, (\breve{v}_{\phi_1}^{(K)})^{-1}\right), \tag{47}$$

$$\pi(\phi_2^{(K)}|M_K) = \prod_{i=1}^{K} f_N\left(\phi_{i,2}|m_{\phi_2}^{(K)}, (\breve{v}_{\phi_2}^{(K)})^{-1}\right); \tag{48}$$

gamma distributions for the coordinates of $\varsigma^{(K)}$ (or, alternatively, the inverse gamma distributions for the variances σ_i^2, $i = 1, 2, \ldots, K$):

$$\pi(\varsigma^{(K)}|M_K) = \prod_{i=1}^{K} f_G\left(\sigma_i^{-2}|\breve{a}^{(K)}, \breve{b}^{(K)}\right); \tag{49}$$

Dirichlet distributions for the (*a priori* independent) rows of the transition matrix:

$$\pi(\eta_1, \eta_2, \ldots, \eta_K|M_K) = \prod_{i=1}^{K} f_{Dir}\left(\eta_i|d_i^{(K)}\right), \tag{50}$$

with $d_i^{(K)} = (d_{i,1}\ d_{i,2}\ \ldots\ d_{i,K})'$ standing for the vector of the hyperparameters.

Note that, quite customarily, equal hyperparameters over the regimes are assumed in (46)-(49). Following the results presented in Propositions 5 and 9, compatible priors under M_1 can be written as

$$\pi(\alpha|M_1) = f_N\left(\alpha|m_\alpha^{(1)}, (\breve{v}_\alpha^{(1)})^{-1}\right), \tag{51}$$

$$\pi(\phi_1|M_1) = f_N\left(\phi_1|m_{\phi_1}^{(1)}, (\breve{v}_{\phi_1}^{(1)})^{-1}\right), \tag{52}$$

$$\pi(\phi_2|M_1) = f_N\left(\phi_2|m_{\phi_2}^{(1)}, (\breve{v}_{\phi_2}^{(1)})^{-1}\right), \tag{53}$$

$$\pi(\sigma^{-2}|M_1) = f_G\left(\sigma^{-2}|\breve{a}^{(1)}, \breve{b}^{(1)}\right), \tag{54}$$

with the hyperparameters related with the ones displayed in (46)-(49) via Formulae (19) and (21) (in the case of the normals), and (31) and (32) (in the case of the gamma distributions).

Since the hyperparameters for each group of the switching parameters under M_K are held equal across the regimes, there are actually two routes available to establish ceherent prior structures. Within the first one, one sets the values of the hyperparameters under the general model first, and then the ones under the single-component model. Within the second approach, one proceeds the other way round. However, should different vaules of the hyperparameters for a given group of the switching parameters under M_K be allowed, then only the first of the two strategies can be followed, with the relevant formulae provided in Propositions 5 and 9.

Eventually, notice that the two: the AR(2) and the MSIAH(K)-AR(2) model, represent the extremes, with the former featuring no switches at all, and the latter, on the other hand, introducting Markovian breaks into all the four coefficients at once: the intercept, the two autoregressive parameters, and the error term's variance. Therefore, the two specifications do not share any common parameters. Naturally, one may be prompted to limit the set of the parameters enabled to switch to include only one, two, or three out of the four, in each case obtaining some "intermediate" specification. Should that be the case, our methodology for establishing compatible priors applies straightforwardly. To deliver some illustrative example, consider an AR(2) model with switches introduced only into the intercept, hereafter denoted as MSI(K)-AR(2) or M_K^*, in short. Obviously, it forms one of all the conceivable "intermediate" specifications, nesting the single-component AR(2) model on the one hand, and being nested within the MSIAH(K)-AR(2) model, on the other. Write $\theta_*^{(K)} = (\alpha^{(K)\prime}\ \phi_1\ \phi_2\ \sigma^{-2}\ \eta')'$ for the vector of M_K^*'s parameters, with $\alpha^{(K)} = (\alpha_1\ \alpha_2\ \dots\ \alpha_K)'$. The prior is structured as

$$\begin{aligned}\pi(\theta_*^{(K)}|M_K^*) = {} & \pi(\alpha^{(K)}|M_K^*)\pi(\phi_1|M_K^*)\pi(\phi_2|M_K^*) \\ & \times \pi(\sigma^{-2}|M_K^*)\pi(\eta|M_K^*)\mathbb{I}_{R_K^*}\{\zeta_K^*(\theta_*^{(K)})\},\end{aligned}$$

where, according to the argumentation presented above, the regularity restriction assumes the form of the one derived for the single-component model: $\zeta_K^*(\theta_*^{(K)}) := \rho(\Phi)$ and $R_K^* = R_1 = [0,1)$. Assuming equal hyperparameters for $\alpha^{(K)}$'s prior, in order to establish such a prior structure under M_K^* that is compatible with that of M_1 we set (46) for $\pi(\alpha^{(K)}|M_K^*)$, and (52)-(54) for $\pi(\phi_1|M_K^*)$, $\pi(\phi_2|M_K^*)$ and $\pi(\sigma^{-2}|M_K^*)$, respectively. Notice that if, in addition to that, the density $\pi(\eta|M_K^*)$ coincides with (50), then the prior structure of M_K^* is also compatible with the one specified under the general model, M_K.

6 Concluding remarks

In this paper we developed the (or rather *some*) concept of compatible prior distributions in univariate mixture and Markov-switching models, including the case where the priors are restrained with some additional restrictions. However, it must be clearly stressed that the problem of priors' compatibility within the considered class of models is yet to be studied in depth, and in that sense the present study's role is much (and, perhaps, merely) of a trigger rather than delivery of conclusive resolution. Conceivably, there are at least several issues that ought to be addressed in that line of research, some of them quite naturally arising from possible allegations one may readily formulate against the approach developed in the current paper. Firstly, suppose, for instance, that the equality constraint on all elements of $\lambda_j^{(K)}$ takes form

$$\lambda_{1,j} = \lambda_{2,j} = \ldots = \lambda_{K,j} \Leftrightarrow \widetilde{\tau}_{2,j} = \widetilde{\tau}_{3,j} = \ldots = \widetilde{\tau}_{K,j} = 1 \Leftrightarrow \widetilde{\tau}_j = 1_{[(K-1)\times 1]},$$

where $\widetilde{\tau}_{i,j} = \lambda_{i,j}/\lambda_{1,j}$. The obvious question is whether the results derived in the current paper would still hold if one were to switch to the above setting (in which the contrasts are defined as ratios rather than differences between the component's individual coordinates), and arguably the correct answer would be "no" (with the phenomenon often referred to as the *Borel-Kolmogorov paradox*; see Dawid and Lauritzen 2001). Deferring any further analysis of the problem, let us note, however, that defining contrasts as simple differences, just as it has been done in our study, is not only natural (specifically, with respect to location parameters and logarithms of scale parameters, as well), but also mathematically favorable, since, firstly, it allows all $\lambda_{i,j}$'s to be fully real-valued (i.e. there is no need to exclude $\lambda_{1,j} = 0$), and, secondly, the Jacobian determinant of the inverse transformation equals 1 (which no longer holds for contrasts defined as ratios).

Secondly, it can be argued that the approach proposed in this study only establishes compatibility between the mixture (or switching) models featuring either K or a single component, whereas it remains unresolved whether it also induces compatibility (in the sense of 1) between a K-component structure and a one with, say, $K-1$ components ($K \geqslant 3$). Again, it appears that the most plausible answer would read "no", but then another issue arises, of how - and if it is possible, at all - to generalize the idea employed in the present article (hinging upon conditioning relevant distributions) to incorporate such cases. On the other hand, there seem to exist in the literature some theoretical results which one could invoke in this context to dispute such a need, at least in practical terms (however theoretically justifiable that need would appear). Namely, Rousseau and Mengersen (2011) prove that - under certain conditions pertaining to the hyperparameters of the Dirichlet prior for the weights of the mixture - the posterior distribution exhibits quite a peculiar feature of emptying the extra (i.e. superfluous) components (once their number set *a priori* turns out to exceed the "correct" number of components so that the mixture model overfits the data). If so, then, conceivably, in practical terms it could do to, first, estimate

a mixture model with a large enough number of components (so as to allow the resulting posterior to reveal the "adequate" number of theirs), and then - if one is still interested in a formal, coherent comparison of the single-component specification and the emerged mixture model - apply the approach presented in the current article to establish compatible priors for both structures. Nevertheless, even if such an approach would seem valid (at least to some extent, though perhaps not entirely elegant, *per se*), the results obtained in the cited paper only pertain to the class of non-dynamic, simple mixture models, therefore excluding Markov-switching specifications (though similar results were attempted to obtain by Gassiat and Rousseau 2014 for hidden Markov models). Hence, it would still seem worth the effort to establish compatibility conditions between mixture (Markov-switching) models featuring different number of components (states, respectively).
Thirdly, in fact, there do exist several strategies of formulating compatible prior distributions, including approaches based upon conditioning (the one adopted in our work), marginalization and the Kullback-Leibler projection (see Dawid and Lauritzen 2001 and Consonni and Veronese 2008 for more details). Obvious lines of further research ensue, including attempts to apply and compare these strategies within the context of models regarded in this study.
Finally, there is still the issue of the impact different strategies of settling compatible priors exert upon the posterior and, additionally, predictive distributions. Such an analysis, of however crucial scientific importance, requires extensive both theoretical and empirical studies, and as such it exceeds by far the scope of the current paper.

Acknowledgements

The publication was prepared with financial support by Foundation for Polish Science (within the START 2014 programme). The author also acknowledges partial support from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

[1] Consonni G. and Veronese P. (2008). Compatibility of prior specifications across linear models. *Statistical Science* 23 (3), 332–353.

[2] Dawid A. and Lauritzen S. (2001). Compatible prior distributions. [in:] George E. (ed.), *Bayesian Methods with Applications to Science, Policy and Official Statistics (Selected Papers from ISBA 2000)*, Monographs of Official Statistics. Eurostat 2001.

[3] Dickey J. (1974). Bayesian alternatives to the F test and the least squares estimate in the normal linear model [:in] Fienberg S. and Zellner A. (eds.), *Studies in Bayesian Econometrics and Statistics*. North-Holland, Amsterdam.

[4] Francq C. and Zakoïan J.-M. (2001). Stationarity of Multivariate Markov-Switching ARMA Models. *Journal of Econometrics* 102, 339–364.

[5] Frühwirth-Schnatter S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer.

[6] Gassiat E. and Rousseau J. (2014). About the posterior distribution in hidden Markov models with unknown number of states. *Bernoulli* 20 (4), 2039–2075.

[7] Jasra A., Holmes C. and Stephens D. (2005). Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science* 20 (1), 50–67.

[8] Krolzig H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Lecture Notes in Economics and Mathematical Systems. New York/Berlin/Heidelberg: Springer.

[9] Marin J.-M., Mengersen K. and Robert C. (2005). Bayesian modelling and inference on mixtures of distributions. [in:] Rao C. and Dey D. (eds.), *Handbook of Statistics: Volume 25*. North-Holland.

[10] Poirier D. (1985). Bayesian hypothesis testing in linear models with continuously induced conjugate priors across hypotheses. [in:] Bernardo J., DeGroot M., Lindley D. and Smith A. (Eds.), *Bayesian Statistics 2*. North-Holland, Amsterdam.

[11] Roeder K. and Wasserman L. (1997). Practical bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association* 92 (439), 894–902.

[12] Rousseau J. and Mengersen K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society, series B* 76, 689–710.

[13] Yao, W. (2012a). Bayesian mixture labeling and clustering. *Communications in Statistics-Theory and Methods* 41 (3), 403–421.

[14] Yao W. (2012b). Model based labeling for mixture models. *Statistics and Computing* 22 (2), 337–347.

A Appendix: Proof of Lemma 2

Firstly, let us briefly recall the assumtions of Lemma 2:

1. In the single-component model, M_K, parameters in δ are *a priori* independent of all the others.

2. In the reparametrized mixture model, $\widetilde{M}_K$, parameters in δ are *a priori* independent of all the others.

3. All prior distributions under consideration are proper.

The thesis of Lemma 2 states that distributions $\pi_{\underline{\theta^{(1)}}}(\theta^{(1)}|M_1)$ and $\pi_{\underline{\theta^{(K)}}}(\theta^{(K)}|M_K)$ (likewise, $\pi_{\underline{\widetilde{\theta}^{(K)}}}(\widetilde{\theta}^{(K)}|\widetilde{M}_K)$) are compatible, i.e.

$$\pi_{\underline{\delta},\underline{\lambda_{1,1}},\underline{\lambda_{1,2}},\ldots,\underline{\lambda_{1,n}}}(\delta,\lambda_{1,1},\lambda_{1,2},\ldots,\lambda_{1,n}|M_1) = \\ = \pi_{\underline{\delta},\underline{\lambda_{1,1}},\underline{\lambda_{1,2}},\ldots,\underline{\lambda_{1,n}}|\underline{\tau_1},\underline{\tau_2},\ldots,\underline{\tau_n}}(\delta,\lambda_{1,1},\ldots,\lambda_{1,n}|\tau_1=\ldots=\tau_n=0_{[(K-1)\times 1]},\widetilde{M}_K), \tag{55}$$

iff the following conditions are satisfied:

$$\pi_{\underline{\delta}}(\delta|M_1) = \pi_{\underline{\delta}}(\delta|\widetilde{M}_K), \tag{56}$$

and, for each $j = 1, 2, \ldots, n$,

$$\pi_{\underline{\lambda_{1,j}}}(\lambda_{1,j}|M_1) = \pi_{\underline{\lambda_{1,j}}|\underline{\tau_j}}(\lambda_{1,j}|\tau_j = 0_{[(K-1)\times 1]},\widetilde{M}_K). \tag{57}$$

For simplicity, in what follows, let us consider the case when $n = 1$.
Necessary condition: if (56) and (57), then (55).

$$\begin{aligned}\pi_{\underline{\delta},\underline{\lambda_{1,1}}}(\delta,\lambda_{1,1}|M_1) =& \pi_{\underline{\delta}}(\delta|M_1)\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1) = \\ &\overset{(56,57)}{=} \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)\pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0_{[(K-1)\times 1]},\widetilde{M}_K) = \\ &\overset{Asmp.\,2}{=} \pi_{\underline{\delta}|\underline{\lambda_{1,1}},\underline{\tau_1}}(\delta|\lambda_{1,1},\tau_1=0,\widetilde{M}_K)\pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0,\widetilde{M}_K) = \\ &= \pi_{\underline{\delta},\underline{\lambda_{1,1}}|\underline{\tau_1}}(\delta,\lambda_{1,1}|\tau_1=0,\widetilde{M}_K).\end{aligned}$$

Sufficient condition: if (55), then (56) and (57).
Factorizing the joint distribution $\pi_{\underline{\delta},\underline{\lambda_{1,1}}}(\delta,\lambda_{1,1}|M_1)$, we obtain:

$$\begin{aligned}\pi_{\underline{\delta},\underline{\lambda_{1,1}}}(\delta,\lambda_{1,1}|M_1) &\overset{(55)}{=} \pi_{\underline{\delta},\underline{\lambda_{1,1}}|\underline{\tau_1}}(\delta,\lambda_{1,1}|\tau_1=0,\widetilde{M}_K) = \\ &\overset{Asmp.\,2}{=} \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)\cdot\pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0,\widetilde{M}_K).\end{aligned} \tag{58}$$

On the other hand,

$$\pi_{\underline{\delta},\underline{\lambda_{1,1}}}(\delta,\lambda_{1,1}|M_1) \overset{Asmp.\,1}{=} \pi_{\underline{\delta}}(\delta|M_1)\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1). \tag{59}$$

Based on (58) and (59), the following equality is established:

$$\pi_{\underline{\delta}}(\delta|M_1)\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1) = \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)\pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0,\widetilde{M}_K). \tag{60}$$

Under Assumption 3, identity (60) indicates the equality between densities of corresponding distributions of δ, once both sides of (60) are integrated with respect to $\lambda_{1,1}$:

$$\int\limits_{\mathbb{R}} \pi_{\underline{\delta}}(\delta|M_1)\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1)d\lambda_{1,1} = \int\limits_{\mathbb{R}} \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)\pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0,\widetilde{M}_K)d\lambda_{1,1}.$$

Hence, $\pi_{\underline{\delta}}(\delta|M_1) = \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)$, which is precisely (56).
On the other hand, integrating (60) with respect to δ, one obtains the equality between densities of corresponding distributions of $\lambda_{1,1}$:

$$\int\limits_{\mathbb{R}} \pi_{\underline{\delta}}(\delta|M_1)\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1)d\delta = \int\limits_{\mathbb{R}} \pi_{\underline{\delta}}(\delta|\widetilde{M}_K)\pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0,\widetilde{M}_K)d\delta.$$

Therefore, $\pi_{\underline{\lambda_{1,1}}}(\lambda_{1,1}|M_1) = \pi_{\underline{\lambda_{1,1}}|\underline{\tau_1}}(\lambda_{1,1}|\tau_1=0,\widetilde{M}_K)$, which is exactly (57).
In conclusion, Equality (60) holds iff conditions (56) and (57) are satisfied.
That Lemma 2 is also true for $n>1$ follows immediately from the prior independence of: $\lambda_{1,1}$, $\lambda_{1,2}$, ..., $\lambda_{1,n}$ in model M_1, and $\lambda_1^{(K)}$, $\lambda_2^{(K)}$, ..., $\lambda_n^{(K)}$ in model M_K. (Additionally, Assumption 3 is needed, as well).

B Appendix: Proof of Proposition 5

Invoking Lemma 3 and performing some simple manipulations, the proof proceeds as follows:

$$\begin{aligned}\pi_{\underline{\lambda_1}}(\lambda_1|M_1) &\propto \prod_{i=1}^{K}\pi_{\underline{\lambda_i}}(\lambda_1|M_K) = \prod_{i=1}^{K} f_N^{(1)}(\lambda_1|m_i^{(K)},v_i^{(K)})\\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{K}\frac{(\lambda_1-m_i^{(K)})^2}{v_i^{(K)}}\right\}\end{aligned}$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{K}\left(\frac{1}{v_i^{(K)}}\lambda_1^2 - 2\frac{m_i^{(K)}}{v_i^{(K)}}\lambda_1\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{K}\frac{1}{v_i^{(K)}}\right)\left(\lambda_1^2 - 2\lambda_1\frac{\sum_{i=1}^{K}\frac{m_i^{(K)}}{v_i^{(K)}}}{\sum_{i=1}^{K}\frac{1}{v_i^{(K)}}}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\left(\sum_{i=1}^{K}\frac{1}{v_i^{(K)}}\right)^{-1}}\left(\lambda_1 - \frac{\sum_{i=1}^{K}\frac{m_i^{(K)}}{v_i^{(K)}}}{\sum_{i=1}^{K}\frac{1}{v_i^{(K)}}}\right)^2\right\}$$

$$\propto f_N^{(1)}(\lambda_1|m^{(1)}, v^{(1)}),$$

with $m^{(1)}$ and $v^{(1)}$ given by (14) and (15), respectively. The proof for the precision-parametrized normal densities follows analogously.

C Appendix: Proof of Proposition 7

The proof is analogous to the one presented for Proposition 5:

$$\pi_{\underline{\lambda_1}}(\lambda_1|M_1) \propto \prod_{i=1}^{K}\pi_{\underline{\lambda_i}}(\lambda_1|M_K) = \prod_{i=1}^{K} f_{IG}(\lambda_1|a_i^{(K)}, b_i^{(K)})$$

$$\propto \left[\prod_{i=1}^{K}(\lambda_1)^{-(a_i^{(K)}+1)}\right]\exp\left\{-\frac{1}{\lambda_1}\sum_{i=1}^{K}\frac{1}{b_i^{(K)}}\right\} =$$

$$= (\lambda_1)^{-\left(\sum_{i=1}^{K}a_i^{(K)}+K-1+1\right)}\exp\left\{-1\Big/\lambda_1\left(\sum_{i=1}^{K}\frac{1}{b_i^{(K)}}\right)^{-1}\right\}$$

$$\propto f_{IG}(\lambda_1|a^{(1)}, b^{(1)}),$$

with $a^{(1)}$ and $b^{(1)}$ given by (22) and (23), respectively.

D Appendix: Proof of Proposition 9

We proceed analogously to the proofs of Propositions 5 and 7:

$$\begin{aligned}\pi_{\underline{\lambda_1}}(\lambda_1|M_1) &\propto \prod_{i=1}^{K}\pi_{\underline{\lambda_i}}(\lambda_1|M_K) = \prod_{i=1}^{K} f_G(\lambda_1|\breve{a}_i^{(K)}, \breve{b}_i^{(K)}) \\ &\propto \left[\prod_{i=1}^{K}(\lambda_1)^{\breve{a}_i^{(K)}-1}\right] \exp\left\{-\lambda_1\sum_{i=1}^{K}\breve{b}_i^{(K)}\right\} = \\ &= (\lambda_1)^{\sum_{i=1}^{K}\breve{a}_i^{(K)}-K+1-1} \exp\left\{-\lambda_1\sum_{i=1}^{K}\breve{b}_i^{(K)}\right\} \\ &\propto f_G(\lambda_1|\breve{a}^{(1)}, \breve{b}^{(1)}),\end{aligned}$$

with $\breve{a}^{(1)}$ and $\breve{b}^{(1)}$ given by (29) and (30), respectively.