Adaptive Subspace Shrinkage with Mixture Functional Horseshoe Priors*

Florian HUBER[a], Matteo IACOPINI[b], and Luca ROSSINI[c]

[a] *University of Salzburg*
[b] *Vrije Universiteit Amsterdam and Tinbergen Institute*
[c] *University of Milan*

September 1, 2022

PRELIMINARY AND INCOMPLETE

Abstract. Striking a balance between model complexity and parsimony is of crucial importance for producing accurate forecasts and structural inference. In this paper, we propose a non-parametric VAR model which relies on splines to approximate the unknown function between the response vector and its lags. This spline-based VAR is highly flexible and potentially overfits the data. To circumvent issues related to overfitting we introduce a prior which shrinks the non-parametric model towards a simpler model. This simpler specification, however, is subject to model uncertainty. We control for this by using a small set of competing linear and non-linear parametric conditional mean models and infer the appropriate submodel towards which we shrink from the data. Our approach allows for deciding whether variable-specific relations are modeled to be completely unknown (and thus a non-parametric technique is most suitable) or whether they exert a simpler, known effect on the endogenous variables. In simulations, we show that our approach accurately detects whether effects are linear or non-linear. We illustrate key model features by estimating a small-scale non-parametric VAR and consider the effects of oil price shocks.

JEL: C11; C14; C32; C51.

KEYWORDS: Bayesian inference; Factor stochastic volatility; Mixture prior; Nonparametric regression; Subspace shrinkage; VAR model.

*Florian Huber gratefully acknowledges financial support from the Austrian Science Fund (FWF, grant no. ZK 35). Matteo Iacopini acknowledges financial support from the EU Horizon 2020 programme under the Marie Skłodowska-Curie scheme (grant agreement no. 887220). We would like to thank .... for helpful comments.

1 Introduction

Economic and financial time series are often subject to substantial structural breaks, changing volatilities and possibly non-linear relations between the endogenous variables and the predictors. These complex patterns typically arise in times of crises such as during the deep recession of 2008/2009 or the Covid-19 pandemic. Especially for the latter episode we have observed outliers far outside the range of past data. Such a situation typically calls for appropriate modeling techniques since linear models fail to capture outliers adequately. This has adverse effects on parameter estimates and ultimately distorts forecasts and impulse responses.

Several recent papers propose non-linear and non-parametric models to capture structural breaks, outliers, heteroskedasticity and possibly changing dynamic relations across variables. Starting with influential work of Primiceri (2005) and Cogley and Sargent (2005), researchers have been working with time-varying parameter (TVP) regressions and VARs. These models are flexible since they allow for changing relationships between the variables in the VAR as well as in the error variances. However, conditional on a specific point in time, they assume a linear relationship between the predictors and the endogenous variables. This is one possible limitation. The second limitation is that these models are tightly parametrized (and thus prone to overfitting). Solutions addressing overfitting can be found in, e.g., Belmonte, Koop, and Korobilis (2014); Bitto and Frühwirth-Schnatter (2019); Huber, Koop, and Onorante (2021); Lopes, McCulloch, and Tsay (2021); Hauzenberger, et al. (2021). However, none of these papers deals with the possible issue of non-linearities conditional on a specific point in time.

Another strand of the literature relies on non-parametric techniques to estimate unknown functional relations in the conditional mean (Kalli and Griffin, 2018; Huber and Rossini, forthcoming; Huber, et al., 2020; Clark, et al., 2021). These models are very flexible, work well if the data features several outliers (for a discussion, see Huber, et al., 2020) and scale well in high dimensions. However, they are also difficult to interpret and shrinkage is often achieved through forcing the underlying response surface towards the space of smoothly varying functions. Since it is not clear how that would impact the underlying set of parameters if the data generating process is linear (as we would expect in tranquil periods) constructing shrinkage priors which shrink towards pre-specific function spaces might be more appropriate (see Shin, Bhattacharya, and Johnson, 2020).

In this paper, we propose a non-parametric VAR which uses splines to approximate the unknown conditional mean function. Instead of using shrinkage priors that force the basis parameters associated with the splines to zero we use the subspace shrinkage prior developed in Shin, Bhattacharya, and Johnson (2020) to shrink the model towards a pre-specified space of known functions. This prior, instead of pushing coefficients to zero, forces the non-parametric model towards a simpler, parametric alternative. As opposed to the original subspace shrinkage prior, we contribute to the literature by addressing one shortcoming: the necessity to select a single subspace towards which we want to shrink our non-parametric model. Our approach solves this by introducing a discrete latent variable which selects an appropriate prior model from a pool of pre-specified models. This is done for each covariate separately. The resulting shrinkage prior allows for variable-specific functional shrinkage, implying that if a given variable exerts a linear (or non-linear) effect on the response, our approach is capable of detecting this accurately.

Another econometric contribution lies in the treatment of the contemporaneous relationships between the shocks to the VAR. Instead of introducing a typical Cholesky-type decomposition of the covariance matrix, we propose a novel non-linear estimator of the covariance matrix. Our covariance specification, again, relies on splines and our modified subspace shrinkage prior to capture non-linearities of an unknown form. Model estimation can be carried out using straightforward Markov chain Monte Carlo (MCMC) techniques and we also provide methods to compute dynamic impulse response functions.

We illustrate our techniques using two applications. First, by means of synthetic data we show that our modified subspace shrinkage prior accurately detects the precise form of non-linearities. But this strongly depends on key parameters of the data generating process and we thus investigate how estimation accuracy changes with features of the DGP. In a second application we use our model to [... complete once the empirical application is clear].

The remainder of this article is organized as follows: Section 2 illustrates the baseline framework within a single equation context. Then, Section 3 describes a novel VAR model with stochastic volatility. Section 4.1 introduces a new class of nonparametric prior distribution and Section 5 presents the Bayesian approach to inference. Section 6 investigates the performance of our method using simulated data. Section 7 shows results of an application to real data on US inflation. Section 8 concludes the article.

2 The univariate model

Our point of departure is a simple non-parametric regression model that posits a unknown and possibly non-linear relationship between a response $y_t \in \mathbb{R}$ and a covariate $x_t \in \mathbb{R}$. The corresponding model is then given by:

$$y_t = f(x_t) + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \exp(h_t)), \tag{1}$$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma_h^2), \tag{2}$$

where $f : \mathbb{R} \to \mathbb{R}$ denotes an unknown function and ϵ_t is a Gaussian shock with time-varying (log) variances h_t. We assume that h_t evolves according to an AR(1) proces with unconditional mean μ_h, persistence ϕ_h and state innovation variance σ_h^2. Conditional on how we estimate f, this model provides flexibility in terms of handling non-linearities in the conditional mean as well as heteroskedasticity in the shocks. The former feature is important of statistical relations between y_t and x_t change over time or are subject to non-linearities whereas the latter feature pays off if interest centers around density and tail forecasting (CITE CCM 2020 paper).

Several options for estimating f are available. For instance, Gaussian processes (CITE CITE), Bayesian additive regression trees (CITE CITE), neural networks or splines can be used to approximating the functional relationship between x_t and y_t. In this paper, we follow Shin, Bhattacharya, and Johnson (2020) and model the unknown function f as spanned by a set of pre-specified B-spline basis functions $\{\phi_j\}_{1 \leq j \leq k_n}$ as follows:

$$f(x) = \sum_{j=1}^{k_n} \beta_j \phi_j(x). \tag{3}$$

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{k_n})'$ denote the vector of basis coefficients and $\Phi = \{\phi_j(x_t)\}_{1 \leq t \leq T, 1 \leq j \leq k_n}$ is a $(T \times k_n)$ matrix of basis functions evaluated at the observed covariates. Stacking all observations and defining $H = \text{diag}(e^{h_1}, \ldots, e^{h_T})$, one can rewrite the model in (1) to (3) as

$$\boldsymbol{y}|\boldsymbol{\beta} \sim \mathcal{N}(\Phi\boldsymbol{\beta}, H). \tag{4}$$

Allowing for K covariates can be straightforwardly achieved by means of a generalized additive model (Hasite and Tibshirani, 1986). This framework models the relationship between a K-dimensional vector of covariates $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{Kt})'$ and a scalar response as the sum of

K univariate functions, where the kth function only dependent on the kth predictor:

$$y_t = \sum_{k=1}^{K} f_k(x_{kt}) + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \exp(h_t)). \tag{5}$$

Using B-splines leads to the multivariate generalization of the model described above:

$$\boldsymbol{y}|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K \sim \mathcal{N}\Big(\sum_{k=1}^{K} \Phi_k \boldsymbol{\beta}_k,\ H\Big), \tag{6}$$

where $\boldsymbol{\Phi}_k$ is a $T \times k_n$-dimensional matrix of basis functions that depend on the kth covariate in $\boldsymbol{x}_t$.

3 Non-parametric VAR models

In macroeconomics and finance, interest often centers around modeling the interaction between several endogenous variables. This is typically achieved through VARs. In this section, we generalize the model outlined in the previous section to the VAR case.

Let us define $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ as the vector of n variables available at time t. We consider the following general VAR model with stochastic volatility (VAR-SV) that assumes a relationship between $\boldsymbol{y}_t$ and its first p lags:

$$\boldsymbol{y}_t = F(\boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p}) + \boldsymbol{\varepsilon}_t, \tag{7}$$

where $F : \mathbb{R}^{np} \to \mathbb{R}^n$. We assume that the shocks $\boldsymbol{\varepsilon}_t$ follow a non-linear factor model:

$$\boldsymbol{\varepsilon}_t = G(\boldsymbol{f}_t) + \boldsymbol{\zeta}_t, \tag{8}$$

with $G : \mathbb{R}^Q \to R^n$ denoting a non-linear function that links $Q(\ll n)$ latent factors $\boldsymbol{f}_t$ to the reduced-form shocks of the VAR and $\boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_t)$ is a vector of measurement errors with a time-varying but diagonal error variance-covariance matrix $\Sigma_t = \text{diag}(e^{h_{1t}}, \ldots, e^{h_{nt}})$.

The factors are assumed to be Gaussian, centered on zero and feature a time-varying covariance matrix $V_t = \text{diag}(e^{v_{1t}}, \ldots, e^{v_{Qt}})$. The logarithms of the diagonal elements of Σ_t and

V_t are assumed to follow stationary AR(1) processes:

$$h_{it} = \mu_{hi} + \phi_{hi}(h_{it-1} - \mu_{hi}) + \eta_{hi,t} \qquad i = 1, \ldots, n,$$
$$v_{qt} = \phi_{vq} v_{qt-1} + \eta_{vq,t} \qquad q = 1, \ldots, Q,$$

where $\eta_{hi,t} \sim \mathcal{N}(0, \sigma^2_{hi})$ and $\eta_{vq,t} \sim \mathcal{N}(0, \sigma^2_{vq})$. This specification implies that conditional on knowing $\{G(\boldsymbol{f}_t)\}_{t=1}^T$, the general VAR reduces to a sequence of independent non-parametric regression models. This speeds up inference enormously. Moreover, assuming that the shocks are effectively driven by a small number of factors reduces the effective number of parameters. Since we remain agnostic on the precise form the function G can take, our model allows for capturing possible non-linearities between shocks in the system and, in addition, controls for heteroskedasticity.

Notice that if $G(\boldsymbol{f}_t) = \Lambda \boldsymbol{f}_t$ is a linear function of the factors, the reduced form covariance matrix of $\boldsymbol{\varepsilon}_t$ can be decomposed as follows:

$$\Omega_t = \Lambda V_t \Lambda' + \Sigma_t. \tag{9}$$

This equation suggests that any contemporaneous relations across shocks are purely driven by the latent factors. Our flexible specification allows for capturing non-linear contemporaneous relations. In macroeconomic and financial applications, this has the advantage that, under suitable structural identification schemes, the model can generate asymmetries in the impact responses to structural shocks. This turns out to be a considerable advantage over other approaches that assume the shocks to be Gaussian with a covariance as in Eq. (9).

The proposed framework nests a wide range of well-known multivariate time series models commonly applied in the literature. *First*, by removing the (nonparametric) spline specification for all the components, the standard VAR with factor SV of Kastner and Huber (2020)is obtained. *Second*, if we model only the conditional mean as an unknown function of the covariates, one gets a nonlinear VAR with linear factor SV. A similar model has been proposed in Clark, et al. (2021). This model allows to capture potentially nonlinear relations between the lagged endogenous variables, while retaining linear unobserved heterogeneous effects. *Third*, assuming a spline specification only for the covariance component results in a linear VAR augmented with a nonlinear factor SV. This model essentially assumes that after filtering out linear effects, the corresponding reduced-form shocks include arbitrary non-linear relations and our model captures them through the function G and the inclusion of latent factors. *Fourth*, if we set the number of factors $Q = n$, all models approach specifications without restrictions on the error covariance matrix.

As will become clear in Section 4.1, as long as the mfHS prior includes the linear specification among the list of possible subspaces, then all these particular cases can be recovered from the general model in Eq. (7) in a completely data-driven manner. This effectively controls for model uncertainty and selection issues.

3.1 Function approximation using B-splines

As discussed in Section 2, we approximate $F = (F_1, \ldots, F_n)'$ and $G = (G_1, \ldots, G_n)'$ using B-splines. The presence of the factor entails equation-by-equation estimation. The model for the i^{th} equation can be written in full-data form as follows:

$$\boldsymbol{y}_{\bullet i} = F_i(\boldsymbol{X}) + G_i(\boldsymbol{f}) + \boldsymbol{\zeta}_{\bullet i}, \quad \boldsymbol{\zeta}_{\bullet i} \sim \mathcal{N}\left(\mathbf{0}, \text{diag}(e^{h_{i1}}, \ldots, e^{h_{iT}})\right), \tag{10}$$

with $\boldsymbol{y}_{\bullet i}$ and $\boldsymbol{\zeta}_{\bullet i}$ denoting the i^{th} column of $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)'$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_T)'$, respectively. The $T \times np$ matrix $\boldsymbol{X}$ has a typical t^{th} row given by $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})'$. Finally, the $T \times Q$ matrix $\boldsymbol{f} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$ stores the latent factors.

We approximate $F_i(\boldsymbol{X})$ and $G_i(\boldsymbol{f})$ through:

$$F_i(\boldsymbol{X}) \approx \sum_{j=1}^{np} \boldsymbol{\Phi}_{ij}^F \boldsymbol{\beta}_{ij}^F, \quad G_i(\boldsymbol{f}) \approx \sum_{q=1}^{Q} \boldsymbol{\Phi}_{iq}^G \boldsymbol{\beta}_{iq}^{(G)}. \tag{11}$$

where $\boldsymbol{\Phi}_{ij}^F, \boldsymbol{\Phi}_{iq}^G$ are $(T \times k_n)$ matrices and $\boldsymbol{\beta}_{ij}^F, \boldsymbol{\beta}_{iq}^G$ are $(k_n \times 1)$ vectors, for each $i = 1, \ldots, n$, $j = 1, \ldots, np$ and $q = 1, \ldots, Q$. This spline-based approximation is highly flexible and allows for unveiling complex patterns in the data. However, this flexibility comes at a cost since overfitting issues naturally arise. We solve these using Bayesian shrinkage priors. The next sub-section discusses our proposed prior setup.

4 The Prior

4.1 Mixture Functional Horseshoe Prior

To set the stage, we first introduce the notation for the functional horseshoe prior (fHS) of Shin, Bhattacharya, and Johnson (2020). Let $\boldsymbol{\Phi}_{ij,0}^{\ell}$ denote a null regressor matrix of size

$(T \times m)$, $m \geq 1$, and define its rank with $d^{\ell}_{ij,0} = \text{rank}(\Phi_{ij,0})$. Denote the projection matrix of $\boldsymbol{\Phi}^{\ell}_{ij,0}$ by $\boldsymbol{Q}^{\ell}_{ij,0} = \boldsymbol{\Phi}^{\ell}_{ij,0}((\boldsymbol{\Phi}^{\ell}_{ij,0})'\boldsymbol{\Phi}^{\ell}_{ij,0})^{-1}(\boldsymbol{\Phi}^{\ell}_{ij,0})'$. Finally, for an $m \times r$ matrix A, with $m > r$ and $\text{rank}(A) = r$, let $\mathfrak{L}(A) = \{A\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^r\}$ denote the column space of A. The fHS prior allows for shrinkage toward a null subspace that is fixed in advance, and defined by means of $\boldsymbol{\Phi}^{\ell}_{ij,0}$. Examples of possible subspaces are the linear (by setting $\boldsymbol{\Phi}^{F}_{ij,0} = \boldsymbol{X}_{\bullet,j}$ or $\boldsymbol{\Phi}^{G}_{ij,0} = \boldsymbol{f}_{\bullet,j}$) or the quadratic (by setting $\boldsymbol{\Phi}^{F}_{ij,0} = (\boldsymbol{X}_{\bullet j}, \boldsymbol{X}^2_{\bullet j})$ or $\boldsymbol{\Phi}^{G}_{ij,0} = (\boldsymbol{f}_{\bullet j}, \boldsymbol{f}^2_{\bullet j}))$ subspace. However, this prior implies that we have to select the subspace towards which we shrink a priori. In this paper we will avoid this and assume that the prior subspace is an unknown quantity which we estimate. This allows us to infer whether the relations encoded by F and G take a linear or non-linear form or whether a given covariate should be excluded from the model.

This is achieved in the following way. The matrix $\boldsymbol{\Phi}^{\ell}_{ij,0}$ is comprised of multiple null regressor matrices, $\boldsymbol{\Phi}^{\ell}_{ij,0} = (\boldsymbol{\Phi}^{\ell}_{ij,01}, \ldots, \boldsymbol{\Phi}^{\ell}_{ij,0L})$, $L \geq 1$. This choice permits shrinkage of the coefficients towards alternative null subspaces. To decide on the specific subspace, we introduce a set of latent allocation variables, $z^{\ell}_{ij} \in \{1, \ldots, L\}$, with prior distribution $z^{\ell}_{ij}|\boldsymbol{p}^{\ell}_{ij} \sim Cat(\boldsymbol{p}^{\ell}_{ij})$, where $Cat(\bullet)$ denotes the categorical distribution and $\boldsymbol{p}^{\ell}_{ij} = (p^{\ell}_{ij,1}, \ldots, p^{\ell}_{ij,L})' \in \Delta^{L-1}$ is a probability vector, with Δ^{L-1} denoting the $(L-1)$-dimensional simplex.

The mixture functional horseshoe prior (mfHS) is given by the following hierarchical prior

$$\begin{aligned}
\pi(\boldsymbol{\beta}^{\ell}_{ij}|\tau^{\ell}_{ij}, z^{\ell}_{ij}) &\propto ((\tau^{\ell}_{ij})^2)^{-\frac{k_n - d^{\ell}_{ij,0z}}{2}} \exp\Big(-\frac{1}{2(\tau^{\ell}_{ij})^2}(\boldsymbol{\beta}^{\ell}_{ij})'(\Phi^{\ell}_{ij})'\Sigma_i^{-1/2}(\boldsymbol{I}_T - \boldsymbol{Q}_{ij,0z})\Sigma_i^{-1/2}\Phi^{\ell}_{ij}\boldsymbol{\beta}^{\ell}_{ij}\Big), \\
\pi(\tau^{\ell}_{ij}) &\propto \frac{((\tau^{\ell}_{ij})^2)^{b-1/2}}{(1 + (\tau^{\ell}_{ij})^2)^{a+b}} \mathbb{I}_{(0,\infty)}(\tau^{\ell}_{ij}), \\
\pi(z^{\ell}_{ij}|\boldsymbol{p}^{\ell}_{ij}) &\propto \prod_{l=1}^{L} (p^{\ell}_{ij,l})^{\mathbb{I}(z^{\ell}_{ij}=l)}, \\
\pi(\boldsymbol{p}^{\ell}_{ij}) &\propto \prod_{l=1}^{L} (p^{\ell}_{ij,l})^{\bar{c}-1},
\end{aligned} \tag{12}$$

where $a, b > 0$ are hyperparameters and $d^{\ell}_{ij,0z}$ and $\boldsymbol{Q}_{ij,0z}$ are the rank and the projection matrix of $\Phi^{\ell}_{ij,0,z^{\ell}_{ij}}$. Note that $\pi(\tau^{\ell}_{ij})$ can be obtained as the prior induced by assuming a $\mathcal{B}eta(a, b)$ prior on $\omega^{\ell}_{ij} = 1/(1 + (\tau^{\ell}_{ij})^2)$.

This prior allows us to investigate, across equations and for each covariate (which includes both $\boldsymbol{X}$ and $\boldsymbol{f}$), whether the effect is of a known non-linear or a linear form. The prior also allows for shrinking the regression effect of a specific covariate towards zero (this is achieved by setting $\boldsymbol{Q}_{ij,0z} = \boldsymbol{0}_{T\times T}$). In case neither of the subspaces fits the data well, our approach also allows for more flexibility by setting τ_{ij}^{ℓ} to a very large value. Intuitively speaking, the parameter τ_{ij}^{ℓ} controls the weight placed on the subspace and the allocation variable governs the subspace towards which we shrink the flexible spline model. For instance, if τ_{ij}^{ℓ} is large, the corresponding prior on $\boldsymbol{\beta}_{ij}^{\ell}$ will be relatively uninformative and the weight placed on the subspace piece will be low. By contrast, if τ_{ij}^{ℓ} is close to zero, substantial prior mass will be allocated to the subspace component and the corresponding conditional mean function closely mimics the behavior of the parametric regression specification embodied in the subspace $\boldsymbol{Q}_{ij,0z}$. Hence, our approach tackles model and specification uncertainty in a very flexible way and requires little input from the researcher.

To provide more intuition on what our prior does, it proves fruitful to focus on a simplified case of the model discussed above. In this simple model, we have a single equation and a single regressor and set $p = n = 1$ so that $\boldsymbol{Y}$ and $\boldsymbol{X}$ are $T-$dimensional vectors. This is the model described in Eq. (1). Hence, we can drop all sub- and superscripts from the quantities described above. In this simple model, one can show that the conditional posterior mean of the regression function given (ω, z) is a convex combination of the B-spline estimator $\boldsymbol{Q}_{\Phi}\boldsymbol{Y}$ and the parametric estimator $\boldsymbol{Q}_{0,z}\boldsymbol{Y}$, where $\boldsymbol{Q}_{\Phi} = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'$. Marginalizing over z yields a convex combination between $\boldsymbol{Q}_{\Phi}\boldsymbol{Y}$ and all $\boldsymbol{Q}_{0,\ell}\boldsymbol{Y}$.

Corollary 1. *Suppose that $\mathfrak{L}(\Phi_{0,\ell}) \not\subset \mathfrak{L}(\Phi)$ for each $\ell = 1, \dots, L$. Then*

$$\mathbb{E}[\Phi\boldsymbol{\beta}|\boldsymbol{Y}, \omega, z] = (1-\omega)\boldsymbol{Q}_{\Phi}\boldsymbol{Y} + \omega\boldsymbol{Q}_{0,z}\boldsymbol{Y}. \tag{13}$$

$$\mathbb{E}[\Phi\boldsymbol{\beta}|\boldsymbol{Y}, \omega, \boldsymbol{p}] = \sum_{\ell=1}^{L} p_{\ell}\Big[(1-\omega)\boldsymbol{Q}_{\Phi}\boldsymbol{Y} + \omega\boldsymbol{Q}_{0,\ell}\boldsymbol{Y}\Big] = \sum_{\ell=1}^{L} \tilde{p}_{1,\ell}\boldsymbol{Q}_{\Phi}\boldsymbol{Y} + \tilde{p}_{0,\ell}\boldsymbol{Q}_{0,\ell}\boldsymbol{Y} \tag{14}$$

$$= \Big(1 - \sum_{\ell=1}^{L}\tilde{p}_{0,\ell}\Big)\boldsymbol{Q}_{\Phi}\boldsymbol{Y} + \sum_{\ell=1}^{L}\tilde{p}_{0,\ell}\boldsymbol{Q}_{0,\ell}\boldsymbol{Y}, \tag{15}$$

where $\tilde{p}_{0,\ell} = p_{\ell}\omega$, $\tilde{p}_{1,\ell} = p_{\ell}(1-\omega)$, and $\sum_{\ell}\tilde{p}_{1,\ell} + \tilde{p}_{0,\ell} = \sum_{\ell} p_{\ell} = 1$.

This result shows that, conditional on the weights $\boldsymbol{p}$ and ω, the posterior fit of our flexible model is a convex combination between the fit of a B-spline model and the fitted values implied by the L different parametric specifications.

4.2 Priors on the remaining model parameters

We follow Kastner and Frühwirth-Schnatter (2014) and assume the following independent prior for the hyper-parameters driving the log-volatility processes:

$$
\begin{aligned}
\mu_{hi} \sim \mathcal{N}(-10,1), \qquad \frac{\phi_{hi}+1}{2} &\sim \mathcal{B}e(50,1.5), \qquad \sigma^2_{hi} \sim \mathcal{G}a(0.5,0.5), \\
\frac{\phi_{vi}+1}{2} &\sim \mathcal{B}e(6,1.5), \qquad \sigma^2_{vi} \sim \mathcal{G}a(0.5,0.5).
\end{aligned} \tag{16}
$$

The choice of the hyperparameters for the prior of the unconditional mean and variance is not influential. The log-volatility processes are restricted to be stationary by imposing a prior distribution on the persistence parameters such that $\phi_{hi}, \phi_{vi} \in (0,1)$. This is done by assuming a Beta prior for the linear transformation $(\phi+1)/2$. The chosen hyperparameters assure a prior mean and standard deviation of 0.94 and 0.05 for ϕ_{hi}, and a smaller prior mean for ϕ_{vi}.

4.3 Selecting the number of factors

Remark 1 (Selection of Q). The proposed mfHS prior allows to make inference on the number of unobserved factors, Q. This is done by including the zero subspace within the "candidate set" of subspaces. Then, at each iteration $m = 1, \dots, M$ of the MCMC algorithm, we count the number of factors that have been allocated to a non-zero subspace. In fact, the allocation variable z^G_{iq} informs about the inclusion (or not) of the qth factor in the regression for the ith endogenous variable. Therefore, we can estimate the number of "active" factors for each endogenous variable as the maximum a posteriori (MAP) of the variable:

$$
N^G_i = \sum_{q=1}^{Q} \mathbb{I}\big(z^{G,(m)}_{iq} \neq 0\big). \tag{17}
$$

Moreover, for each variable i and factor q we can compute the posterior inclusion probability (PIP), that is the posterior probability that factor q is relevant to variable i, as:

$$
PIP^G_{iq} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{I}\big(z^{G,(m)}_{iq} \neq 0\big). \tag{18}
$$

5 Posterior sampling

The paths of the log-volatilities and the corresponding parameters are sampled as in Kastner and Frühwirth-Schnatter (2014) using the R-package `stochvol` (Kastner, 2016).

The main steps of the MCMC algorithm are reported below (for $i = 1, \ldots, n$, $j = 1, \ldots, n$, $l = 1, \ldots, q$, and $t = 1, \ldots, T$) and we refer to the supplement for further details:

1. sample the spline coefficients and related hyper-parameters ($t = 1, \ldots, T$, $i = 1, \ldots, n$, and $j = 1, \ldots, n$):
   (a) sample $\boldsymbol{\beta}^F_{ij} | \boldsymbol{Y}, \tau^F_{ij}, z^F_{ij}, \boldsymbol{h}_i, \boldsymbol{\beta}^G, \boldsymbol{f}$ from the Gaussian full conditional $\mathcal{N}(\boldsymbol{\beta}^F_{ij} | \bar{\beta}^F_{ij}, \bar{\Sigma}^F_{\beta,ij})$;
   (b) sample $\tau^F_{ij} | \boldsymbol{Y}, \boldsymbol{\beta}^F_{ij}, \boldsymbol{h}_i, z^F_{ij}$ from the full conditional $p(\tau^F_{ij} | \boldsymbol{Y}, \boldsymbol{\beta}^F_{ij}, \boldsymbol{h}_i, z^F_{ij})$ using a slice sampler;
   (c) sample $z^F_{ij} | \boldsymbol{\beta}^F_{ij}, \boldsymbol{p}^F_{ij}$ from the categorical full conditional $Cat(z^F_{ij} | \bar{\boldsymbol{p}}^F_{ij})$;
   (d) sample $\boldsymbol{p}^F_{ij} | z^F_{ij}$ from the Dirichlet full conditional $\mathcal{D}ir(\boldsymbol{p}^F_{ij} | \bar{\boldsymbol{c}}^F_{ij})$;
2. sample the factors, the spline coefficients, and related hyper-parameters ($t = 1, \ldots, T$, $i = 1, \ldots, n$, and $q = 1, \ldots, Q$):
   (a) sample the factors $\boldsymbol{f}_t | \boldsymbol{y}_t, \boldsymbol{\beta}^F, \boldsymbol{\beta}^G, \boldsymbol{h}_t, \boldsymbol{v}_t$ using a Metropolis-Hastings step, with proposal given by the posterior distribution of a linear factor stochastic volatility specification;
   (b) sample $\boldsymbol{\beta}^G_{iq} | \boldsymbol{Y}, \tau^G_{iq}, z^G_{iq}, \boldsymbol{v}_i, \boldsymbol{f}$ from the Gaussian full conditional $\mathcal{N}(\boldsymbol{\beta}^G_{iq} | \bar{\beta}^G_{iq}, \bar{\Sigma}^G_{\beta,iq})$;
   (c) sample $\tau^G_{iq} | \boldsymbol{f}, \boldsymbol{\beta}^G_{iq}, \boldsymbol{v}_i, z^G_{iq}$ from the full conditional $p(\tau^G_{iq} | \boldsymbol{F}, \boldsymbol{\beta}^G_{iq}, \boldsymbol{v}_i, z^G_{iq})$ using a slice sampler;
   (d) sample $z^G_{iq} | \boldsymbol{\beta}^G_{iq}, \boldsymbol{p}^G_{iq}$ from the categorical full conditional $Cat(z^G_{iq} | \bar{\boldsymbol{p}}^G_{iq})$;
   (e) sample $\boldsymbol{p}^G_{iq} | z^G_{iq}$ from the Dirichlet full conditional $\mathcal{D}ir(\boldsymbol{p}^G_{iq} | \bar{\boldsymbol{c}}^G_{iq})$;
3. sample the history of the log-volatilities and related hyper-parameters, given $\boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{\beta}, \Lambda$, using the R-package `stochvol` (Kastner, 2016).

6 Evidence using Artificial Data

6.1 Univariate model

In this section, we test the ability of the model proposed in Section 3 to XXXXX

To assess the classification performance of the model, Fig. 2 shows the posterior distribution of the allocation variable, z_j. We find that, in settings where the type of relationship between the response and the covariate is included in the collection of subspaces $\{\Phi_l\}$, then our model shrinks towards the correct function. Conversely, when the true relationship is not represented in the collection of subspaces (e.g., sinusoidal case), then the model assigns higher weight to the nonparametric spline part (as reported by ω).

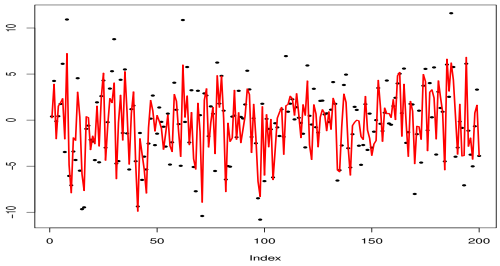

Figure 1: Observed (dots) and fitted values (solid line). Univariate model, with $K = 9$ covariates.

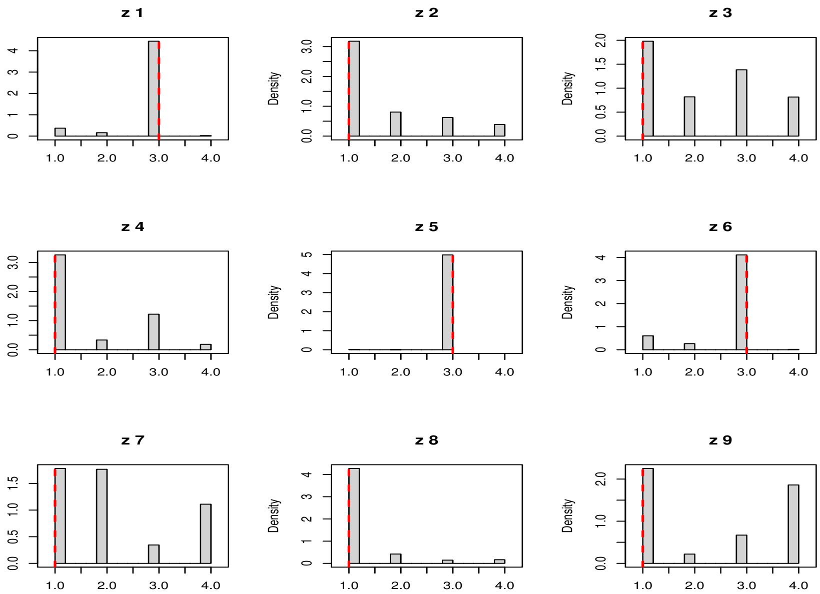

Figure 2: Posterior distribution (histogram) and real value (dashed red line) of the allocation variable z_k, $k = 1, \ldots, K$.

		$\sigma = 0.001$	$\sigma = 0.1$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 3$
linear	$K = 3$	0.883	0.914	0.945	0.922	0.827
		(0.152)	(0.139)	(0.120)	(0.133)	(0.198)
	$K = 5$	0.833	0.822	0.844	0.842	0.711
		(0.144)	(0.165)	(0.151)	(0.173)	(0.205)
quadratic	$K = 3$	0.690	0.688	0.662	0.715	0.777
		(0.201)	(0.186)	(0.203)	(0.232)	(0.191)
	$K = 5$	0.630	0.657	0.653	0.690	0.676
		(0.176)	(0.168)	(0.170)	(0.171)	(0.181)

Table 1: F1 Score computed over different simulated scenario

6.2 VAR model

Synthetic data have been generated from a nonlinear, homoskedastic, stationary VAR(1) model, where the relationship between the endogenous variables and their lagged values is one of the following: (i) linear, (ii) quadratic, (iii) sinusoidal, or (iv) zero (i.e., variable $y_{j,t-1}$ has no impact on $y_{i,t}$). We consider a medium-sized model with $n = 20$ and fix the noise covariance to $\Sigma = 0.5^2 I_n$. The proposed VAR model is estimated assuming $q = 2$ factors and independent, stationary stochastic volatility processes for the idiosyncratic component of the noise variance. Overall, this results in the estimation of a miss-specified model.

Figure 3 compares the simulated data and the fitted values, providing evidence of the ability of our method to correctly recover the observed paths in the data, also in presence of rapid changes and non-stationary behaviors.

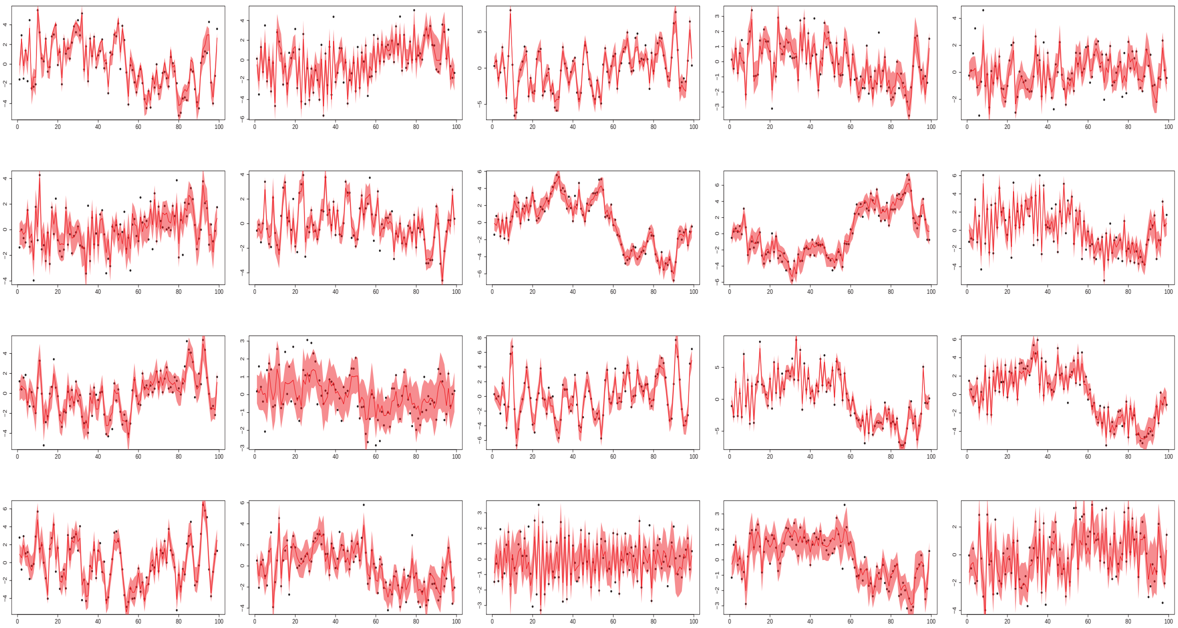

Figure 3: Observed (dots), fitted values (solid line), and 95% credible intervals (color shade). VAR with factor SV model, with $n = 20$ endogenous variables and true variance $\sigma_y^2 = 0.50 I_n$.

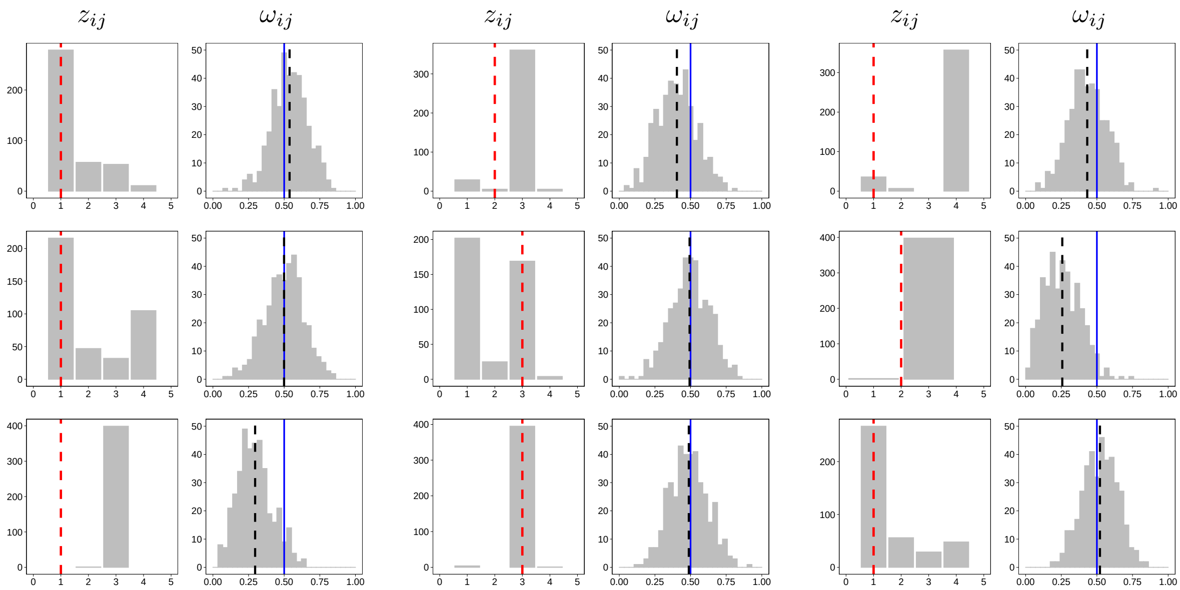

Figure 4: Left: posterior distribution (histogram) and true value (red, dashed line) of some of the allocation variable z_{ij}. Right: posterior distribution (histogram), posterior mean (black, dashed line), and 0.5 line (blue, solid line) of the parameter ω_{ij}. VAR with factor SV model, with $n = 20$ endogenous variables and true variance $\sigma_y^2 = 0.50 I_n$.

7 Real Data Application to US inflation (univariate)

As common practice in VAR models with factor stochastic volatility, we investigate the propagation of exogenous shocks within the system by considering the effects of a change in the unobserved factor, $\mathbf{f}_t$. Specifically, to obtain an impulse response function we simulate a shock to the k-th factor, $f_{k,t}$, and compute the associated change to the other variables. Instantaneously, the transmission is driven by the respective factor loadings, whereas the lagged effects are given by the recursive structure of the system.

8 Conclusions

We have proposed an extension of VAR models that accounts for nonlinear and sparse effects, idiosyncratic stochastic volatility, and unobserved heterogeneity. The flexibility is achieved by means of a novel class of nonparametric prior distributions, the mixture functional horseshoe prior (mfHS), which generalizes the subspace shrinkage priors of Shin, Bhattacharya, and Johnson (2020) in two directions. First, we go beyond the univariate regression and define a prior of interest for investigating multivariate time series models. Second, we introduce a mixture to allow (possibly) different shrinkage for each response-covariate variables pair.

Moreover, we capture potentially nonlinear and sparse unobserved heterogeneity effects by specifying the proposed mfHS prior for the latent factors, thus extending the linear factor stochastic volatility model.

References

Belmonte, Miguel AG, Gary Koop, and Dimitris Korobilis (2014), "Hierarchical shrinkage in time-varying parameter models," *Journal of Forecasting*, 33, 80–94.

Bitto, Angela, and Sylvia Frühwirth-Schnatter (2019), "Achieving shrinkage in a time-varying parameter model framework," *Journal of Econometrics*, 210, 75–97.

Clark, Todd E, Florian Huber, Gary Koop, Massimiliano Marcellino, and Michael Pfarrhofer (2021), "Investigating growth at risk using a multi-country non-parametric quantile factor model," *arXiv preprint arXiv:2110.03411*.

Cogley, Timothy, and Thomas J Sargent (2005), "Drifts and volatilities: monetary policies and outcomes in the post wwii us," *Review of Economic dynamics*, 8, 262–302.

Hasite, Trevor, and Robert Tibshirani (1986), "Generalized additive models," *Statistical Science*, 1, 297–318.

Hauzenberger, Niko, Florian Huber, Gary Koop, and Luca Onorante (2021), "Fast and flexible bayesian inference in time-varying parameter regression models," *Journal of Business & Economic Statistics*, 1–15.

Huber, Florian, Gary Koop, and Luca Onorante (2021), "Inducing sparsity and shrinkage in time-varying parameter models," *Journal of Business & Economic Statistics*, 39, 669–683.

Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (2020), "Nowcasting in a pandemic using non-parametric mixed frequency vars," *Journal of Econometrics*.

Huber, Florian, and Luca Rossini (forthcoming), "Inference in bayesian additive vector autoregressive tree models," *Annals of Applied Statistics*.

Kalli, Maria, and Jim E Griffin (2018), "Bayesian nonparametric vector autoregressive models," *Journal of econometrics*, 203, 267–282.

Kastner, Gregor (2016), "Dealing with stochastic volatility in time series using the R package stochvol," *Journal of Statistical Software*, 69, 1–30.

Kastner, Gregor, and Sylvia Frühwirth-Schnatter (2014), "Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models," *Computational Statistics & Data Analysis*, 76, 408–423.

Kastner, Gregor, and Florian Huber (2020), "Sparse bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, 39, 1142–1165.

Lopes, Hedibert F, Robert E McCulloch, and Ruey S Tsay (2021), "Parsimony inducing priors for large scale state–space models," *Journal of Econometrics*.

Primiceri, Giorgio E (2005), "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72, 821–852.

Shin, Minsuk, Anirban Bhattacharya, and Valen E Johnson (2020), "Functional horseshoe priors for subspace shrinkage," *Journal of the American Statistical Association*, 115, 1784–1797.

A Posterior sampling

The posterior sampling for the extended model samples from the full conditional distributions below.

Sample the coefficient vector $\boldsymbol{\beta}_j^{(i)}$ The spline coefficient vectors, $\boldsymbol{\beta}_j^{(i)}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n$, are sampled from the Gaussian full conditional distribution

$$\boldsymbol{\beta}_j^{(i)} | \boldsymbol{y}_i, H_i, \omega_{ij}, z_{ij} \sim \mathcal{N}(\bar{\boldsymbol{\beta}}_{z,ij}, \bar{\Sigma}_{z,ij}) \tag{19}$$

where $\bar{\Sigma}_{z,ij} = (\Phi'_{ij} H_i^{-1} \Phi_{ij} + \frac{\omega_{ij}}{1-\omega_{ij}} \Phi'_{ij}(I - Q_{0,z_{ij}})\Phi_{ij})^{-1}$ and $\bar{\boldsymbol{\beta}}_{z,ij} = \bar{\Sigma}_{z,ij} \Phi'_{ij} \boldsymbol{y}_i$.

Sample the auxiliary parameter τ_{ij} The full conditional distribution of the auxiliary parameter τ_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, n$, is

$$\begin{aligned} \tau_{ij} | \boldsymbol{y}, \boldsymbol{\beta}_j^{(i)}, H_i, z_{ij} \propto (\tau_{ij}^2)^{-(k_n - d_{0,z_{ij}})/2 + b - 1/2} (1 + \tau_{ij}^2)^{-(a+b)} \\ \cdot \exp\big(- \boldsymbol{\beta}_j^{(i)\prime} \Phi'_{ij} H_i^{-1/2} (I - Q_{0,z_{ij}}) H_i^{-1/2} \Phi_{ij} \boldsymbol{\beta}_j^{(i)} / 2 \big). \end{aligned}$$

We use a slice sampler based on the reparametrization $\eta_{ij} = 1/\tau_{ij}^2$. The conditional posterior for η_{ij} is

$$\begin{aligned} \eta_{ij} | \boldsymbol{y}, \boldsymbol{\beta}_j^{(i)}, H_i, z_{ij} \propto \eta_{ij}^{a + (k_n - d_{0,ij})/2 - 1} (1 + \eta_{ij})^{-(a+b)} \\ \cdot \exp\big(- \boldsymbol{\beta}_j^{(i)\prime} \Phi'_{ij} H_i^{-1/2} (I - Q_{0,z_ij}) H_i^{-1/2} \Phi_{ij} \boldsymbol{\beta}_j^{(i)} / 2 \big). \end{aligned}$$

Let $T\mathcal{G}a(x|a, b, S) \propto \mathcal{G}a(x|a, b)\mathbb{I}_S(x)$ denote the Gamma distribution truncated on S. We proceed as follows

- set $r_{ij} = (1 + \eta_{ij})^{-(a+b)}$. Sample the slice variable

$$u_{ij} | \eta_{ij} \sim \mathcal{U}(0, r_{ij}), \tag{20}$$

- set $r_{ij}^* = u_{ij}^{-(a+b)^{-1} - 1}$ and $S = (0, r_{ij}^*)$. Sample η_{ij} from

$$\eta_{ij} | \boldsymbol{y}, \boldsymbol{\beta}_j^{(i)}, H_i, u_{ij}, z_{ij} \sim T\mathcal{G}a(a_{0,ij}, b_{0,ij}, S_{ij}), \tag{21}$$

where $a_{0,ij} = a + (k_n - d_{0,z_{ij}})/2$ and $b_{0,ij} = \boldsymbol{\beta}_j^{(i)\prime} \Phi'_{ij} H_i^{-1/2} (I - Q_{0,z_{ij}}) H_i^{-1/2} \Phi_{ij} \boldsymbol{\beta}_j^{(i)} / 2$.

Finally, set $\tau_{ij}^2 = \eta_{ij}^{-1}$.

Sample the allocation variable z_{ij}

$$P(z_{ij} = \ell | \boldsymbol{\beta}_j^{(i)}, H_i, \omega_{ij}, \boldsymbol{p}_{ij}) \propto p_{\ell,ij} P(\boldsymbol{\beta}_j^{(i)} | \omega_{ij}, z_{ij} = \ell) \tag{22}$$

Sample the mixing probability vector $\boldsymbol{p}_{ij}$ Let $n_{z,ij} = (\mathbb{I}_{\{1\}}(z_{ij}), \dots, \mathbb{I}_{\{L\}}(z_{ij}))'$. The full conditional distribution of the probability vector $\boldsymbol{p}_{ij}$ is the Dirichlet distribution:

$$\boldsymbol{p}_{ij} | z_{ij} \sim \mathcal{D}ir(\bar{c} + n_{z,ij}). \tag{23}$$

Sample the path of the stochastic volatility of the observations, H, and related hyperparameters The full history of the individual (log-)stochastic volatilities of the observations, $\{h_{it}\}_t$, $i = 1, \dots, n$, is sampled along with the hyper-parameters driving the dynamics, $(\mu_{h,i}, \phi_{h,i}, \sigma_{h,i}^2)$, using the R `stochvol` package.

Sample the latent factors $\boldsymbol{f}_t$ Let $\Sigma_t = \text{diag}(e^{-h_{1,t}/2}, \dots, e^{-h_{n,t}/2})$, $\tilde{\boldsymbol{y}}_t = \Sigma_t(\boldsymbol{y}_t - \Phi\boldsymbol{\beta})$, $\tilde{\Lambda}_t = \Sigma_t \Lambda$, and $V_t = \text{diag}(e^{-v_{1,t}/2}, \dots, e^{-v_{q,t}/2})$. The full conditional distribution of the latent factors $\boldsymbol{f}_t$, $t = 1, \dots, T$, is the Gaussian distribution

$$\boldsymbol{f}_t | \boldsymbol{y}_t, \boldsymbol{\beta}, H_t, V_t, \Lambda \sim \mathcal{N}(\bar{\mu}_{f,t}, \bar{\Sigma}_{f,t}),$$

where $\bar{\Sigma}_{f,t} = (\tilde{\Lambda}_t' \tilde{\Lambda}_t + V_t)^{-1}$ and $\bar{\mu}_{f,t} = \bar{\Sigma}_{f,t} \Lambda' \tilde{\boldsymbol{y}}_t$.

Sample the factor loadings Λ Let $\Sigma_i = \text{diag}(e^{-h_{i,1}/2}, \dots, e^{-h_{i,T}/2})$, $\tilde{\boldsymbol{y}}_i = \Sigma_i(\boldsymbol{y}_i - \sum_{j=1}^n \Phi_{ij} \boldsymbol{\beta}_j^{(i)})$ and $\tilde{\boldsymbol{f}}_i = \Sigma_i F$. The full conditional distribution of each row of the factor loadings $\Lambda_{i,\bullet}$, $i = 1, \dots, n$, is the Gaussian distribution

$$\Lambda_{i,\bullet} | \boldsymbol{y}, \boldsymbol{\beta}, H, F \sim \mathcal{N}(\bar{\mu}_{\lambda,i}, \bar{\Sigma}_{\lambda,i}),$$

where $\bar{\Sigma}_{\lambda,i} = (\tilde{\boldsymbol{f}}_i' \tilde{\boldsymbol{f}}_i + I_q)^{-1}$ and $\bar{\mu}_{\lambda,i} = \bar{\Sigma}_{\lambda,i} \tilde{\boldsymbol{f}}_i' \tilde{\boldsymbol{y}}_i$.

Sample the path of the stochastic volatility of the latent factors, V, and related hyperparameters The full history of the individual (log-)stochastic volatilities of the latent factors, $\{v_{lt}\}_t$, $l = 1, \dots, 1$, is sampled along with the hyper-parameters driving the dynamics,

$(\phi_{v,l}, \sigma^2_{v,l})$, using the R `stochvol` package. We refer to Kastner and Frühwirth-Schnatter (2014) for further details. In particular, the log-volatilities are joint sampled all without a loop. The sampler adapts the ancillarity-sufficiency interweaving strategy, which exploits a reparametrization of the log-volatility process to transfer the level of log-variance μ and/or its volatility σ^2 from the state process to the observation process. These parameters are sampled twice in each iteration, one in the centered and one in non-centered parametrizations, and the overall scheme reduces the correlation of MCMC draws. In practice, in the centered parametrization all the parameters are drawn using a MH step, whereas in the non-centered one (μ, σ) are jointly sampled from a bivariate normal full conditional and ϕ via a MH step with normal proposal.