WORKING PAPER

5/2020

Flexible Fat-tailed Vector Autoregression

Sune Karlsson and Stepan Mazur

Statistics

ISSN 1403-0586

https://www.oru.se/institutioner/handelshogskolan/forskning/working-papers/
Örebro University School of Business
701 82 Örebro
SWEDEN

Flexible Fat-tailed Vector Autoregression

Sune Karlsson* Stepan Mazur†

Örebro University School of Business

April 24, 2020

Abstract

We propose a general class of multivariate fat-tailed distributions which includes the normal, t and Laplace distributions as special cases as well as their mixture. Full conditional posterior distributions for the Bayesian VAR-model are derived and used to construct a MCMC-sampler for the joint posterior distribution. The framework allows for selection of a specific special case as the distribution for the error terms in the VAR if the evidence in the data is strong while at the same time allowing for considerable flexibility and more general distributions than offered by any of the special cases. As fat tails can also be a sign of conditional heteroskedasticity we also extend the model to allow for stochastic volatility. The performance is evaluated using simulated data and the utility of the general model specification is demonstrated in applications to macroeconomics.

JEL-codes: C11, C15, C16, C32, C52

Keywords: Scale mixture of normals; Elliptically contoured distribution; Mixture distributions; Stochastic volatility; Markov Chain Monte Carlo

*sune.karlsson@oru.se

†stepan.mazur@oru.se

1 Introduction

Bayesian VAR-models have become the de facto standard model in much of applied macroeconomics and also with many applications in finance. Departing from the basic constant parameter, constant variance, normally distributed VARs of Kadiyala and Karlsson (1997) and Sims and Zha (1998) the model has been generalized in several directions in order to better cope with the challenges of applied work and the vagaries of data. Notable and useful extensions of the basic VAR include allowing for time varying parameters (Doan et al., 1984; Highfield, 1987; Sims, 1993; Cogley and Sargent, 2002) and/or stochastic volatility (Uhlig, 1997; Cogley and Sargent, 2005; Primiceri, 2005; Chan and Eisenstat, 2018).

The normality assumption has however, with few exceptions, remained unchallenged despite the widespread recognition that data is frequently non-normal. The limited literature on BVARs with non-normal error distributions include Ni and Sun (2005) who study a VAR with multivariate-t distributed error terms and derives posterior distributions and MCMC-samplers for a range of prior distributions on the parameters in the VAR. Panagiotelis and Smith (2008) consider a BVAR with a multivariate skew-t distributions and use this to forecast electricity prices. Lee et al. (2016) consider VAR-models with multivariate-t and Laplace distributions for the error terms. The focus in the paper is on the selection of shrinkage parameter for the normal-Wishart or independent normal Wishart prior. The posterior modes are taken as the parameter estimates thereby avoiding computationally intensive MCMC. Chiu et al. (2017) propose a VAR model with fat tails and stochastic volatility where the "structural" errors, $\boldsymbol{\varepsilon}_t$, have univariate t-distributions. With a triangular identification scheme the reduced form errors, $\boldsymbol{u}_t = A\boldsymbol{\varepsilon}_t$, are then a weighted sum of the independent t-distributed structural errors.

Common to these papers is that the non-normal distribution is obtained as a scale mixture of a normal distribution, i.e. $\boldsymbol{u}_t \sim \mathcal{N}(0, \xi_t \boldsymbol{\Sigma})$ for some non-negative random variable ξ_t. One important implication of this is that the likelihood is normal and heteroskedastic conditional on ξ_t and full conditional posteriors for the parameters in the VAR-model are readily available. It is thus relatively straightforward to implement a MCMC-sampler for the joint posterior distribution with ξ_t as a latent variable.

While a t-distribution or Laplace can be an improvement on the normality assumption it is, in many cases, not obvious that this is the best choice. The class of scale mixtures of normals (or more generally elliptically contoured distributions) is quite rich and include a range of distributions as special cases, e.g. Cauchy and the slash distribution in addition to the t and Laplace. Instead of focusing on a specific special case as the earlier literature we propose a more general specification of the distribution of the error term while staying in a framework that leads to conditional normality and feasible MCMC-samplers. We do this by using a mixture of scale mixtures (or a MSM distribution) where the weights in the mixture are estimated from the data. This will allow the data to speak about the shape of the distribution and, if the evidence in the data is strong, the selection of a known special case (including the normal distribution) as the error distribution. At the same time our specification is flexible enough to include more general shapes and intermediate cases.

The paper is structured as follows. First, in Section 2, we introduce the reader to the family of elliptically contoured distributions and give the details of our new MSM distribution. In Section 3, we deliver the full conditional posterior distributions for the Bayesian VAR-model and construct a MCMC-sampler for the joint posterior distribution.

Section 4 extends all the results to the Bayesian VAR-model with stochastic volatility. Section 5 discuss estimation of the marginal likelihood and model choice. The properties of the model and methods are illustrated using simulated data in Section 6, while Section 7 provides applications to macroeconomic data. Section 8 concludes.

2 A flexible fat-tailed distribution

We take the class of elliptically contoured distributions as the starting point when constructing a flexible fat-tailed distribution for the error terms in the VAR-model. Recall that a m-dimensional random vector $\boldsymbol{x}$ is elliptically contoured distributed if its density function exists and is given by

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} h\left((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^m$, $\boldsymbol{\Psi}$ is a $m \times m$ positive definite covariance matrix, $h : [0,\infty) \to [0,\infty)$ (see Gupta et al. (2013) for details). This distribution will be denoted by $\mathcal{E}_m(\boldsymbol{\mu},\boldsymbol{\Psi},h)$. Alternatively, the distribution can be characterized through its stochastic representation. If $\boldsymbol{x} \sim \mathcal{E}_m(\boldsymbol{\mu},\boldsymbol{\Psi},h)$, then it holds that

$$\boldsymbol{x} \stackrel{d}{=} \boldsymbol{\mu} + \zeta_0 \boldsymbol{\Psi}^{1/2}\boldsymbol{v},$$

where ζ_0 is a nonnegative random variable, $\boldsymbol{v}$ is a m-dimensional random vector uniformly distributed on the unit sphere in $\mathbb{R}^m$, ζ_0 and $\boldsymbol{v}$ are independent and $\stackrel{d}{=}$ denotes equality in distribution.

Except in special cases the stochastic representation can also be written as a scale mixture of normals,

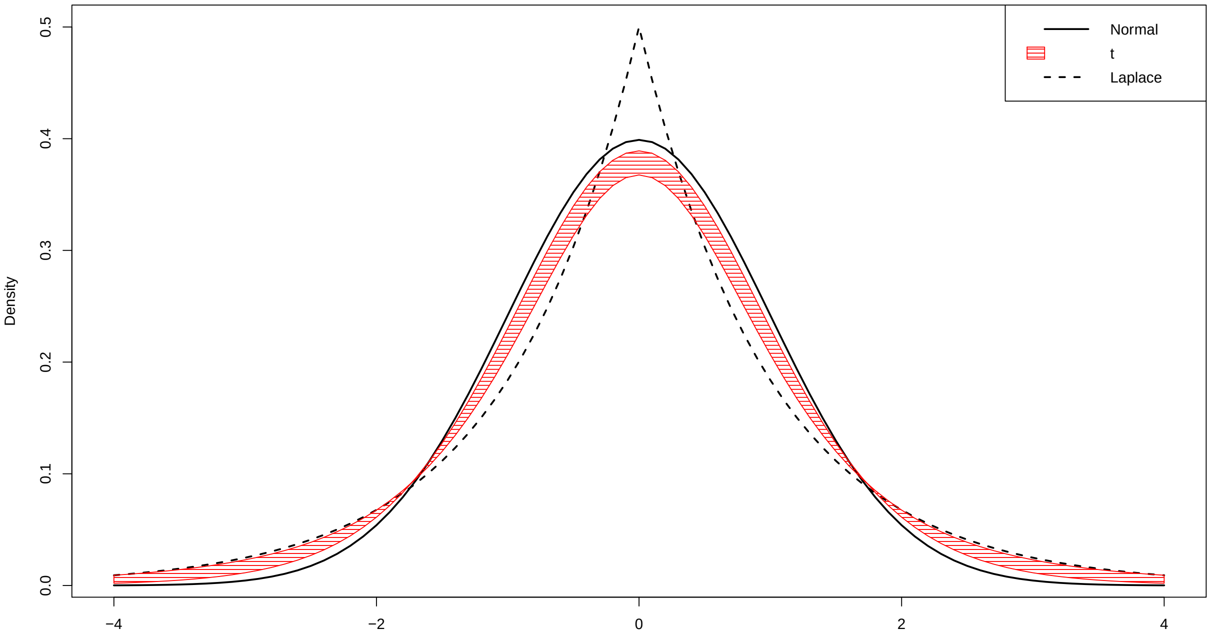
$$\boldsymbol{x} \stackrel{d}{=} \boldsymbol{\mu} + \zeta \boldsymbol{\Psi}^{1/2}\boldsymbol{u},$$

with $\zeta = \zeta_0/\|\boldsymbol{u}\|$, $\boldsymbol{u} \sim \mathcal{N}_m(\boldsymbol{0},\boldsymbol{I}_m)$, and ζ_0, $\|\boldsymbol{u}\|$, $\boldsymbol{u}$ are mutually independently distributed. For simplicity we will only consider distributions that can be obtained as a scale mixture of normals. This still leaves us with a very rich family of distributions with, for example, the multivariate normal, t, and Laplace distributions as special cases, each corresponding to a specific mixing distribution for ζ. Where we depart from earlier work - and this is at the heart of our contribution - is that we do not simply replace the normality assumption with one of the other special cases. Instead we let the mixing distribution itself be a discrete mixture of distributions. This allows the data to speak about the shape of the distribution and can effectively select one of the special cases if the posterior distribution of the weights is concentrated on a single component in the mixture. At the same time it does not force a specific choice on the data and can accommodate more complex shapes of the distribution by allocating posterior weight to several components.

It turns out to be more convenient to work with the distribution of $\xi = \zeta^2$. Collecting the parameters, θ_i, of the components as well as the weights of the mixture in $\boldsymbol{\theta}$ we write the mixing distribution as the mixture

$$f_\xi\left(\xi_t|\boldsymbol{\theta}\right) = \sum_{i=1}^{K} w_i f_{\xi,i}\left(\xi_t|\theta_i\right) \tag{1}$$

Figure 1 Distributional shapes covered by the mixture



Shaded area show distributional shapes covered by the t-distribution with $3 \leq \tau \leq 10$.

leading to a mixture of scale mixtures, or MSM distribution for short. For the special cases above we have

$$f_{\xi,i}\left(\xi_t|\theta_i\right) = \begin{cases} \delta(\xi_t = 1) & \text{point mass at 1 for a multivariate normal component} \\ \mathcal{IG}\left(\tau/2, (\tau-2)/2\right) & \text{inverse gamma for a multivariate } t \text{ component with } \tau \text{ df} \\ \mathcal{G}\left(1,1\right) & \text{gamma for a multivariate Laplace component} \end{cases} \tag{2}$$

For the normal and Laplace components all moments exist and the variance is given by $\boldsymbol{\Psi}$ but the distributions have distinctly different shapes. Using the standard definition, the multivariate-t component would have variance $\frac{\tau}{\tau-2}\boldsymbol{\Psi}$. To facilitate interpretation as well as moves between the components in the MCMC-sampler we reparameterize (rescale) to obtain a common variance $\boldsymbol{\Psi}$ for all components. This in turn leads to the nonstandard mixing distribution in (2). Figure 1 shows the different distributional shapes covered by the components in the mixture. As the t-distribution converge to a normal as $\tau \to \infty$ we limit this parameter both below and above in order achieve separation between the distributions.

3 Prior and posterior distributions

To fix notation, write the VAR-model as

$$\begin{aligned} \boldsymbol{y}_t' &= \sum_{i=1}^{p} \boldsymbol{y}_{t-i}' \boldsymbol{A}_i + \boldsymbol{x}_t' \boldsymbol{C} + \boldsymbol{u}_t' \\ &= \boldsymbol{z}_t' \boldsymbol{\Gamma} + \boldsymbol{u}_t', \end{aligned}$$

where $\boldsymbol{x}_t$ is a vector of d deterministic variables, $\boldsymbol{z}_t' = (\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}', \boldsymbol{x}_t')$ is a $k = mp + d$ dimensional vector, $\boldsymbol{\Gamma}' = (\boldsymbol{A}_1', \ldots, \boldsymbol{A}_p', \boldsymbol{C}')$ is a $k \times m$ variate matrix of constants.

Letting $\boldsymbol{u}_t|\xi_t \sim \mathcal{N}_m(\boldsymbol{0}, \xi_t\boldsymbol{\Psi})$ where ξ_t has the mixture distribution (1) gives the MSM-VAR. Stacking the data for T observations we can rewrite the VAR-model as

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Gamma} + \boldsymbol{U},$$

where $\boldsymbol{Y}$ and $\boldsymbol{U}$ are $T \times m$ matrices and $\boldsymbol{Z}$ is a $T \times k$ matrix. The distribution of $\boldsymbol{Y}$ conditional on $\{\xi_t\}_{t=1}^{T}$ is thus matricvariate normal

$$\boldsymbol{Y}|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Xi} \sim \mathcal{N}_{T,m}(\boldsymbol{Z}\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Xi}),$$

where $\boldsymbol{\Xi} = \mathrm{diag}(\xi_1, \xi_2, \ldots, \xi_T)$ for $\xi_i = \zeta_i^2$ and $\mathcal{N}_{T,m}$ denotes the matricvariate normal distribution for a $T \times m$ matrix (see Karlsson, 2013, for notation).

It remains to specify the prior distributions for $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\theta}$ (the parameters of the mixture). For $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ the two most common forms of prior distributions are the normal-Wishart prior,

$$\begin{aligned} \boldsymbol{\Gamma}|\boldsymbol{\Psi} &\sim \mathcal{N}_{k,m}\left(\underline{\boldsymbol{\Gamma}}, \boldsymbol{\Psi}, \underline{\boldsymbol{\Omega}}\right) \\ \boldsymbol{\Psi} &\sim \mathcal{IW}_m\left(\underline{\boldsymbol{S}}, \underline{\nu}\right), \end{aligned} \tag{3}$$

and independent normal Wishart prior,

$$\begin{aligned} \boldsymbol{\gamma} &\sim \mathcal{N}\left(\underline{\boldsymbol{\gamma}}, \underline{\boldsymbol{V}}\right) \\ \boldsymbol{\Psi} &\sim \mathcal{IW}_m\left(\underline{\boldsymbol{S}}, \underline{\nu}\right), \end{aligned} \tag{4}$$

where $\mathcal{IW}$ denotes the inverse Wishart distribution and $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{\Gamma})$.

The prior for the parameters of the mixing distribution is specified as independent of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ with a Dirichlet prior for the weights, w_i, in the mixture (1),

$$\boldsymbol{w} \sim \mathcal{D}(\underline{\boldsymbol{\alpha}}),$$

and a truncated exponential prior for τ in the multivariate-t component,

$$\pi(\tau) \propto \underline{\lambda}_\tau \exp(-\tau\underline{\lambda}_\tau), \tau_{min} \le \tau \le \tau_{max}$$

where τ_{min} reflects prior notions about the existence of moments and τ_{max} ensures separation between the distributions.[1] Finally, $\boldsymbol{\theta} = (\boldsymbol{w}, \tau)$ collects the parameters of the mixture distribution for ξ_t.

With prior independence between $(\boldsymbol{\Gamma}, \boldsymbol{\Psi})$ and $\boldsymbol{\theta}$ the posterior for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ is independent of $\boldsymbol{\theta}$ once we condition on the latent variable ξ_t. We can thus work with the conditional likelihood

$$\begin{aligned} \mathcal{L}(\boldsymbol{Y}|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Xi}) &= (2\pi)^{-\frac{mT}{2}}|\boldsymbol{\Xi}|^{-\frac{m}{2}}|\boldsymbol{\Psi}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Xi}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\Gamma})\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\Gamma})'\right]\right\} \\ &\propto |\boldsymbol{\Xi}|^{-\frac{m}{2}}|\boldsymbol{\Psi}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Psi}^{-1}\boldsymbol{S}_{\boldsymbol{\Xi}}\right]\right\} \\ &\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Psi}^{-1}(\boldsymbol{\Gamma} - \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}})'\boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Z}(\boldsymbol{\Gamma} - \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}})\right]\right\}. \end{aligned} \tag{5}$$

[1] See Figure 1

where $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}} = (\boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Y}$, the GLS estimate, and $\boldsymbol{S}_{\boldsymbol{\Xi}} = (\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}})'\boldsymbol{\Xi}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}})$.

Standard calculations then yield the full conditional posteriors for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ as (see Karlsson, 2013, for details)

$$\begin{aligned} \boldsymbol{\Gamma}|\boldsymbol{\Psi},\boldsymbol{\Xi},\boldsymbol{Y} &\sim \mathcal{N}_{k,m}\left(\overline{\boldsymbol{\Gamma}}_{\Xi},\boldsymbol{\Psi},\overline{\boldsymbol{\Omega}}_{\Xi}\right), && (6)\\ \boldsymbol{\Psi}|\boldsymbol{\Xi},\boldsymbol{Y} &\sim \mathcal{IW}_m(\overline{\nu},\overline{\boldsymbol{S}}_{\Xi}), && (7) \end{aligned}$$

with

$$\begin{aligned} \overline{\boldsymbol{\Omega}}_{\Xi}^{-1} &= \underline{\boldsymbol{\Omega}}^{-1} + \boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Z}, && (8)\\ \overline{\boldsymbol{\Gamma}}_{\Xi} &= \overline{\boldsymbol{\Omega}}_{\Xi}(\underline{\boldsymbol{\Omega}}^{-1}\underline{\boldsymbol{\Gamma}} + \boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Z}\widehat{\boldsymbol{\Gamma}}_{\Xi}) = \overline{\boldsymbol{\Omega}}_{\Xi}\left(\underline{\boldsymbol{\Omega}}^{-1}\underline{\boldsymbol{\Gamma}} + \boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Y}\right),\\ \overline{\boldsymbol{S}}_{\Xi} &= \underline{\boldsymbol{S}} + \underline{\boldsymbol{\Gamma}}'\underline{\boldsymbol{\Omega}}^{-1}\underline{\boldsymbol{\Gamma}} + \boldsymbol{Y}'\boldsymbol{\Xi}^{-1}\boldsymbol{Y} - \overline{\boldsymbol{\Gamma}}_{\Xi}'\overline{\boldsymbol{\Omega}}_{\Xi}^{-1}\overline{\boldsymbol{\Gamma}}_{\Xi}\\ &= \underline{\boldsymbol{S}} + \boldsymbol{S}_{\boldsymbol{\Xi}} + \left(\underline{\boldsymbol{\Gamma}} - \widehat{\boldsymbol{\Gamma}}_{\Xi}\right)'\left[\underline{\boldsymbol{\Omega}} + \left(\boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Z}\right)^{-1}\right]^{-1}\left(\underline{\boldsymbol{\Gamma}} - \widehat{\boldsymbol{\Gamma}}_{\Xi}\right),\\ \overline{\nu} &= T + \underline{\nu} \end{aligned}$$

for the normal-Wishart prior (3) and

$$\begin{aligned} \gamma|\boldsymbol{\Psi},\boldsymbol{\Xi},\boldsymbol{Y} &\sim \mathcal{N}\left(\overline{\gamma}_{\Xi},\overline{\boldsymbol{V}}_{\Xi}\right), && (9)\\ \boldsymbol{\Psi}|\gamma,\boldsymbol{\Xi},\boldsymbol{Y} &\sim \mathcal{IW}_m\left(\overline{\boldsymbol{S}},\overline{\nu}\right), && (10) \end{aligned}$$

with

$$\begin{aligned} \overline{\boldsymbol{V}}_{\Xi}^{-1} &= \underline{\boldsymbol{V}}^{-1} + \boldsymbol{\Psi}^{-1}\otimes\boldsymbol{Z}^{T}\boldsymbol{\Xi}^{-1}\boldsymbol{Z}, && (11)\\ \overline{\gamma}_{\Xi} &= \overline{\boldsymbol{V}}_{\Xi}\left[\underline{\boldsymbol{V}}^{-1}\underline{\gamma} + \left(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\boldsymbol{Z}\right)\widehat{\gamma}_{\Xi}\right] = \overline{\boldsymbol{V}}_{\Xi}\left[\underline{\boldsymbol{V}}^{-1}\underline{\gamma} + \left(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{Z}'\boldsymbol{\Xi}^{-1}\right)\boldsymbol{y}\right],\\ \overline{\boldsymbol{S}}_{\Xi} &= \underline{\boldsymbol{S}} + \left(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\Gamma}\right)'\boldsymbol{\Xi}^{-1}\left(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\Gamma}\right),\\ \overline{\nu} &= T + \underline{\nu} \end{aligned}$$

for the independent normal Wishart prior (4) and $\widehat{\gamma}_{\Xi} = \mathrm{vec}(\widehat{\boldsymbol{\Gamma}}_{\Xi})$ the GLS estimate.

The posterior distribution for several other common prior distributions for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ are available as special cases of the results above. The posterior corresponding to the diffuse Jeffreys' prior $\pi\left(\boldsymbol{\Gamma},\boldsymbol{\Psi}\right) \propto |\boldsymbol{\Psi}|^{-(m+1)/2}$ is obtained by setting $\underline{\boldsymbol{\Omega}}^{-1} = \mathbf{O}$, $\underline{\boldsymbol{S}} = \mathbf{O}$ and $\underline{\nu} = -k$ in (8). Similarly the posterior for the Normal-diffuse prior, $\boldsymbol{\gamma} \sim \mathcal{N}\left(\underline{\gamma},\underline{\boldsymbol{V}}\right)$ and $\pi\left(\boldsymbol{\Psi}\right) \propto |\boldsymbol{\Psi}|^{-(m+1)/2}$ is obtained by setting $\underline{\boldsymbol{S}} = \mathbf{O}$ and $\underline{\nu} = 0$ in (11).

The posterior distribution of the latent ξ_t conditional on $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ follows from the conditional likelihood (5) and the prior independence of $\boldsymbol{\theta}$ and is given by

$$\begin{aligned} p(\boldsymbol{\Xi}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\theta},\boldsymbol{Y}) \propto |\boldsymbol{\Xi}|^{-m/2}\exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\Gamma}\right)'\boldsymbol{\Xi}^{-1}\left(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{\Gamma}\right)\right]\right\}\\ \times\prod_{t=1}^{T} f_{\xi}\left(\xi_t|\boldsymbol{\theta}\right). \end{aligned}$$

That the full conditional posterior $p(\boldsymbol{\Xi}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\theta},\boldsymbol{Y})$ has the same form irrespective of the prior for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ as long as it is independent of $\boldsymbol{\theta}$ is noteworthy. In addition, it is easily verified that the elements of $\boldsymbol{\Xi}$, ξ_t, are independent conditional on $\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\theta}$ and $\boldsymbol{Y}$. Let

$\boldsymbol{u}_t$ be row t of $\boldsymbol{U} = \boldsymbol{Y} - \boldsymbol{Z\Gamma}$ and q_t the quadratic form $q_t = \boldsymbol{u}_t \boldsymbol{\Psi}^{-1} \boldsymbol{u}_t'$, we then have the full conditional posterior for ξ_t as

$$p(\xi_t | \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \boldsymbol{Y}) \propto \xi_t^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t}\right) \sum_{i=1}^{K} w_i f_{\xi,i}\left(\xi_t | \theta_i\right),$$

the product of an inverse gamma distribution with shape parameter $m/2 - 1$ and rate parameter $q_t/2$, $\mathcal{IG}\left(m/2-1, q_t/2\right)$, and the density of ξ.

To facilitate MCMC we introduce a latent indicator variable $s_t = 1, \ldots, K$, with $p(s_t = j) = w_j$ where ξ_t is drawn from component j if $s_t = j$. Conditionally on s_t we then have

$$p(\xi_t | \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, s_t, \boldsymbol{Y}) \propto \xi_t^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t}\right) f_{\xi,s_t}\left(\xi_t | \theta_{s_t}\right)$$

and the joint conditional distribution of ξ_t and s_t is given by

$$\begin{aligned} p(\xi_t, s_t | \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \boldsymbol{Y}) &= p(\xi_t | \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, s_t, \boldsymbol{Y}) p(s_t | \boldsymbol{w}) \\ &\propto \xi_t^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t}\right) f_{\xi,s_t}\left(\xi_t | \theta_{s_t}\right) \prod_{j=1}^{K} w_j^{\delta_{t,j}}, \end{aligned} \quad (12)$$

where $\delta_{t,j} = 1$ if $s_t = j$ and zero otherwise.

Explicit expressions for the conditional posterior distributions $p(\xi_t | \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, s_t, \boldsymbol{Y})$ and $p(s_t | \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \xi_t, \boldsymbol{Y})$ are straightforward but not needed as we will generate draws from the joint full conditional posterior (12) using a Metopolis-Hastings step for better mixing of the Markov chain.[2] The Metropolis-Hastings step for a joint draw of s_t and ξ_t works as follows. Given the current values (s_t, ξ_t), generate a proposal s_t^* from $p(s_t | \boldsymbol{w})$ and ξ_t^* from $f_{\xi,s_t^*}\left(\xi_t | \theta_{s_t^*}\right)$. The acceptance probability is then

$$\begin{aligned} \alpha &= \min\left(1, \frac{(\xi_t^*)^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t^*}\right) f_{\xi,s_t^*}\left(\xi_t^* | \theta_{s_t^*}\right) w_{s_t^*}}{\xi_t^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t}\right) f_{\xi,s_t}\left(\xi_t | \theta_{s_t}\right) w_{s_t}} \frac{f_{\xi,s_t}\left(\xi_t | \theta_{s_t}\right) w_{s_t}}{f_{\xi,s_t^*}\left(\xi_t^* | \theta_{s_t^*}\right) w_{s_t^*}}\right) \quad (13) \\ &= \min\left(1, \frac{(\xi_t^*)^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t^*}\right)}{\xi_t^{-m/2} \exp\left(-\frac{q_t/2}{\xi_t}\right)}\right). \end{aligned}$$

Accept (s_t^*, ξ_t^*) with probability α, otherwise retain (s_t, ξ_t).

[2]The basic issue is that a Gibbs sampler sampling from the full conditional posteriors for s_t and ξ_t will not be irreducible as $P(s_t = \text{normal} | \xi_t \neq 1) = 0$ and $p(\xi_t = 1 | s_t \neq \text{normal}) = 0$.

Finally, the full conditional posterior for $\boldsymbol{\theta}$ is

$$\begin{aligned}
p(\boldsymbol{w},\tau,\eta|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\xi},\boldsymbol{s},\boldsymbol{Y}) &\propto \prod_{t=1}^{T} p(\xi_t,s_t|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\theta},\boldsymbol{Y})p_w(\boldsymbol{w})p_\tau(\tau) \\
&\propto \prod_{t=1}^{T} \xi_t^{-m/2}\exp\left(-\frac{q_t/2}{\xi_t}\right) f_{\xi,s_t}\left(\xi_t|\theta_{s_t}\right)\prod_{j=1}^{K} w_j^{\delta_{t,j}} \\
&\times p_w(\boldsymbol{w})p_\tau(\tau) \\
&\propto \prod_{j=1}^{K} w_j^{n_j}\prod_{t:s_t=j} f_{\xi,j}\left(\xi_t|\theta_j\right)p_w(\boldsymbol{w})p_\tau(\tau) \\
&\propto \prod_{j=1}^{K} w_j^{n_j}\prod_{t:s_t=j} f_{\xi,j}\left(\xi_t|\theta_j\right)\times\exp(-\tau\underline{\lambda}_\tau)\prod_{j=1}^{K} w_j^{\underline{\alpha}_j-1}
\end{aligned}$$

where $n_j=\sum_t \delta_{t,j}$ is the number of observations for component j of the mixture.

The full conditional posteriors for $\boldsymbol{w}$, τ and η are then straightforward. For the weights/probabilities in the mixture we have a Dirichlet posterior,

$$\boldsymbol{w}|\boldsymbol{s}\sim\mathcal{D}(\overline{\boldsymbol{\alpha}}), \tag{14}$$

for $\overline{\alpha}_j=\underline{\alpha}_j+n_j$. The full conditional posterior of the degrees of freedom in the t-distributed component is given by

$$\begin{aligned}
\tau|\boldsymbol{s},\boldsymbol{\xi} &\propto \exp(-\tau\underline{\lambda}_\tau)\prod_{t:s_t=j}\frac{(\tau-2)^{\tau/2}}{\Gamma(\tau/2)}\xi_t^{-(\tau/2+1)}\exp\left(-\frac{\tau-2}{2\xi_t}\right) \\
&\propto (\tau-2)^{n_j\tau/2}\left[\Gamma(\tau/2)\right]^{-n_j}\exp(-\bar{\lambda}_\tau\tau),\tau_{min}\leq\tau\leq\tau_{max}
\end{aligned} \tag{15}$$

where $s_t=j$ corresponds to the inverse gamma component of the mixture and $\bar{\lambda}_\tau=\underline{\lambda}_\tau+\frac{1}{2}\sum_{t:s_t=j}(\ln\xi_t+\xi_t^{-1})$.

For the conditional posterior (15) for τ we use a Metropolis-Hastings step with a random walk proposal $\tau^*\sim\mathcal{N}(\tau,c_\tau)$ given the current value τ with c_τ chosen to achieve a reasonable acceptance probability. The acceptance probability simplifies to

$$\alpha=\min\left(1,\frac{(\tau^*-2)^{n_j\tau^*/2}[\Gamma(\tau^*/2)]^{-n_j}\exp(-\bar{\lambda}_\tau\tau^*)I(\tau_{min}\leq\tau^*\leq\tau_{max})}{(\tau-2)^{n_j\tau/2}[\Gamma(\tau/2)]^{-n_j}\exp(-\bar{\lambda}_\tau\tau)}\right). \tag{16}$$

Algorithm 1 summarizes the MCMC algorithm for generating draws from the joint posterior distribution.

4 Stochastic volatility

Empirically, fat tails in the error distribution can be difficult to distinguish from time-varying volatility. To investigate this we extend the model to allow for stochastic volatility in addition to the flexible fat-tailed mixture distribution for the error terms for a MSM-SV-VAR.

In particular, we let the variance-covariance matrix be time varying with $\boldsymbol{\Psi}_t=\boldsymbol{P}^{-1}\boldsymbol{\Lambda}_t\boldsymbol{P}^{-1\prime}$ where $\boldsymbol{P}$ is a lower triangular matrix with ones on the diagonal and $\boldsymbol{\Lambda}_t=\text{diag}\left(\exp(h_{1,t}),\ldots,\exp(h_{m,t})\right)$.

Algorithm 1 MCMC algorithm for the MSM-VAR

Initialize $\boldsymbol{\Xi}^{(0)} = \boldsymbol{I}$
Initialize $w_i^{(0)} = 1/K$, $i = 1, \ldots, K$
Initialize $\tau^{(0)} = \tau_{min} + 10$
if independent Normal-Wishart prior **then**
  Initialize $\boldsymbol{\Gamma}^{(0)} = \hat{\boldsymbol{\Gamma}}$
end if
for $j = 1$ **to** $B + R$ **do**
  if Normal-Wishart prior **then**
    Draw $\boldsymbol{\Psi}^{(j)}$ from $\boldsymbol{\Psi}|\boldsymbol{\Xi}^{(j-1)}, \boldsymbol{Y}$ in (7)
    Draw $\boldsymbol{\Gamma}^{(j)}$ from $\boldsymbol{\Gamma}|\boldsymbol{\Psi}^{(j)}, \boldsymbol{\Xi}^{(j-1)}, \boldsymbol{Y}$ in (6)
  else if independent Normal-Wishart prior **then**
    Draw $\boldsymbol{\Psi}^{(j)}$ from $\boldsymbol{\Psi}|\boldsymbol{\Gamma}^{(j-1)}, \boldsymbol{\Xi}^{(j-1)}, \boldsymbol{Y}$ in (10)
    Draw $\boldsymbol{\Gamma}^{(j)}$ from $\boldsymbol{\Gamma}|\boldsymbol{\Psi}^{(j)}, \boldsymbol{\Xi}^{(j-1)}, \boldsymbol{Y}$ in (9)
  end if
  for $t = 1$ **to** T **do**
    Draw $(s_t, \xi_t)^{(j)}$ from $(s_t, \xi_t)|\boldsymbol{\Gamma}^{(j)}, \boldsymbol{\Psi}^{(j)}, \boldsymbol{w}^{(j-1)}, \tau^{(j-1)}, \eta^{(j-1)}, \boldsymbol{Y}$ in (12) using the Metropolis-Hastings step with acceptance probability (13)
  end for
  Draw $\boldsymbol{w}^{(j)}$ from $\boldsymbol{w}|\boldsymbol{s}^{(j)}$ in (14)
  For the multivariate t component draw $\tau^{(j)}$ from $\tau|\boldsymbol{s}^{(j)}, \boldsymbol{\xi}^{(j)}$ in (15) using the Metropolis-Hastings step with acceptance probability (16)
end for
Discard the first B draws as burn-in.

For the log-volatilities, $h_{i,t}$, we follow Frühwirth-Schnatter and Wagner (2010) and specify them as a stationary process,

$$h_{i,t} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \breve{u}_{i,t}, \quad \breve{u}_{i,t} \sim \mathcal{N}(0, \breve{\sigma}_i^2), \quad i = 1, 2, \ldots, m, \tag{17}$$

with the initial state drawn from the unconditional distribution,

$$h_{i,0}|\mu_i, \phi_i, \breve{\sigma}_i \sim \mathcal{N}\left(\mu_i, \breve{\sigma}_i^2/(1 - \phi_i^2)\right).$$

The volatility parameters are given independent priors, normal for the constant terms,

$$\mu_i \sim \mathcal{N}(\mu_\mu, \sigma_\mu^2), \tag{18}$$

and a rescaled Beta distribution for ϕ_i that enforces stationarity,

$$(\phi_i + 1)/2 \sim \mathcal{B}(\alpha_\phi, \beta_\phi). \tag{19}$$

For the variance of the log-volatility, $\breve{\sigma}_i$, we again follow Frühwirth-Schnatter and Wagner (2010) and specify a Gamma prior rather than the usual inverser Gamma as this does not preclude a zero variance and constant volatility. Specifically,

$$\breve{\sigma}_i^2 \sim C_{\breve{\sigma}_i} \cdot \chi_1^2 = \mathcal{G}\left(1/2, 1/2C_{\breve{\sigma}_i}\right). \tag{20}$$

Stacking the rows below the diagonal of $\boldsymbol{P}$ into $\boldsymbol{p}$, e.g. $\boldsymbol{p} = (p_{2,1}, p_{3,1}, p_{3,2}, \ldots, p_{m,m-1})'$, we specify a multivariate normal prior for $\boldsymbol{p}$,

$$\boldsymbol{p} \sim \mathcal{N}(\underline{\boldsymbol{p}}, \underline{\boldsymbol{V}}_p). \tag{21}$$

For the remaining parameters, $\boldsymbol{\Gamma}$ and $\boldsymbol{\theta} = (\boldsymbol{w}, \tau)$ the priors are the same as in the previous section with the independent Normal prior (4) for $\boldsymbol{\gamma}$. The only difference is that we, for computational convenience, reparametrize the model in terms of $\boldsymbol{\beta} = \text{vec}(\boldsymbol{\Gamma}')$ with the corresponding normal prior for $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} \sim \mathcal{N}(\underline{\boldsymbol{\beta}}, \underline{\boldsymbol{V}}_\beta).$$

4.1 Full conditional posteriors

$\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are conditionally independent of the volatility parameters $(\mu_i, \phi_i, \breve{\sigma}_i^2)$ given ξ_t and $\boldsymbol{\Psi}_t$. With $\boldsymbol{u}_t|\xi_t, \boldsymbol{\Psi}_t \sim \mathcal{N}_m(\boldsymbol{0}, \xi_t\boldsymbol{\Psi}_t)$ the conditional likelihood function is given by

$$\begin{aligned}
\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\Gamma}, \boldsymbol{\Psi}_t, \boldsymbol{\Xi}) &= (2\pi)^{-\frac{mT}{2}} |\boldsymbol{\Xi}|^{-\frac{m}{2}} \prod_{t=1}^{T} |\boldsymbol{\Psi}_t|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} (\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma}) (\xi_t\boldsymbol{\Psi}_t)^{-1} (\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma})'\right\} \\
&= (2\pi)^{-\frac{mT}{2}} |\boldsymbol{\Upsilon}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (\widetilde{\boldsymbol{y}} - \boldsymbol{W}\boldsymbol{\beta})' \boldsymbol{\Upsilon}^{-1} (\widetilde{\boldsymbol{y}} - \boldsymbol{W}\boldsymbol{\beta})\right\}
\end{aligned}$$

with $\widetilde{\boldsymbol{y}} = \text{vec}(\boldsymbol{Y}')$, $\boldsymbol{\Upsilon} = \text{diag}(\xi_1\boldsymbol{\Psi}_1, \ldots, \xi_T\boldsymbol{\Psi}_T)$ and $\boldsymbol{W} = \boldsymbol{Z} \otimes \boldsymbol{I}_m$.

After some algebra it is easy to see that the posterior distribution for $\boldsymbol{\beta}$ is conditionally normal, that is,

$$\boldsymbol{\beta}|\boldsymbol{\Psi}_t, \boldsymbol{\Xi}, \boldsymbol{Y} \sim \mathcal{N}\left(\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{V}}_\beta\right), \tag{22}$$

with

$$\begin{aligned}
\overline{\boldsymbol{V}}_\beta^{-1} &= \underline{\boldsymbol{V}}_\beta^{-1} + \boldsymbol{W}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{W} \\
&= \underline{\boldsymbol{V}}_\beta^{-1} + (\boldsymbol{Z} \otimes \boldsymbol{P})' \widetilde{\boldsymbol{\Upsilon}}^{-1} (\boldsymbol{Z} \otimes \boldsymbol{P}), \\
\overline{\boldsymbol{\beta}} &= \overline{\boldsymbol{V}}_\beta \left[\underline{\boldsymbol{V}}_\beta^{-1}\underline{\boldsymbol{\beta}} + \boldsymbol{W}'\boldsymbol{\Upsilon}^{-1}\widetilde{\boldsymbol{y}}\right] \\
&= \overline{\boldsymbol{V}}_\beta \left[\underline{\boldsymbol{V}}_\beta^{-1}\underline{\boldsymbol{\beta}} + (\boldsymbol{Z} \otimes \boldsymbol{P})' \widetilde{\boldsymbol{\Upsilon}}^{-1} \text{vec}(\boldsymbol{P}\boldsymbol{Y}')\right],
\end{aligned}$$

where we have used that $\boldsymbol{\Upsilon} = (\boldsymbol{I}_T \otimes \boldsymbol{P}^{-1})\widetilde{\boldsymbol{\Upsilon}}(\boldsymbol{I}_T \otimes \boldsymbol{P}^{-1'})$ for $\widetilde{\boldsymbol{\Upsilon}}$ the diagonal matrix

$$\begin{aligned}
\widetilde{\boldsymbol{\Upsilon}} &= \text{diag}(\xi_t\boldsymbol{\Lambda}_t) \\
&= \text{diag}\left[\xi_1 \exp(h_{1,1}), \ldots, \xi_1 \exp(h_{m,1}), \ldots, \xi_T \exp(h_{1,T}), \ldots, \xi_T \exp(h_{m,T})\right].
\end{aligned}$$

The full conditional posteriors for $\boldsymbol{w}$ and τ only depend on the latent states s_t and ξ_t and are unaffected by the introduction of stochastic volatility and thus given by (14) and (15).

The conditional posterior distribution of $\boldsymbol{\Xi}$ is expressed as

$$\begin{aligned}
\boldsymbol{\Xi}|\boldsymbol{\Gamma}, \boldsymbol{\Psi}_t, \boldsymbol{\theta}, \boldsymbol{Y} \propto |\boldsymbol{\Xi}|^{-\frac{m}{2}} \exp&\left\{-\frac{1}{2}\sum_{t=1}^{T} (\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma}) (\xi_t\boldsymbol{\Psi}_t)^{-1} (\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma})'\right\} \\
&\times \prod_{t=1}^{T} f_\xi(\xi_t|\boldsymbol{\theta}).
\end{aligned}$$

Moreover, it holds that

$$\xi_t|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \boldsymbol{Y} \propto \xi_t^{-m/2} \exp\left(-\frac{\breve{q}_t/2}{\xi_t}\right) \sum_{i=1}^{K} w_i f_{\xi,i}\left(\xi_t|\theta_i\right),$$

where $\breve{q}_t = \left(\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma}\right) \boldsymbol{\Psi}_t^{-1} \left(\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma}\right)' = \left[\left(\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma}\right) \boldsymbol{P}'\right] \boldsymbol{\Lambda}_t^{-1} \left[\left(\boldsymbol{y}_t' - \boldsymbol{z}_t'\boldsymbol{\Gamma}\right) \boldsymbol{P}'\right]'$. The states s_t and ξ_t can thus be sampled using the Metropolis-Hasting step (13) with q_t replaced by $\breve{q}_t$.

Let $\boldsymbol{\varepsilon}_t = \xi_t^{-1/2}\boldsymbol{P}\boldsymbol{u}_t$, $\varepsilon_{i,t}$ is then mean zero with stochastic volatility and log volatilities given by (17). We can thus sample from the posterior distributions of the latent log-variances $h_{i,t}$ and the parameters $(\mu_i, \phi_i, \breve{\sigma}_i)$ using the MCMC method of Kastner and Frühwirth-Schnatter (2014) as implemented in the `stochvol` `R` package (Kastner, 2016).

For drawing $\boldsymbol{P}$, we consider the following system

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} + u_{1,t}p_{21} \\ u_{3,t} + u_{2,t}p_{32} + u_{1,t}p_{31} \\ \vdots \end{bmatrix} = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \vdots \end{bmatrix},$$

where $\epsilon_{i,t} \sim \mathcal{N}(0, \xi_t \exp(h_{i,t}))$, $i = 1, \ldots, m$. We thus have $m-1$ heteroskedastic regressions.

$$\begin{aligned} u_{2,t} &= -u_{1,t}p_{21} + \epsilon_{2,t}, \\ u_{3,t} &= -u_{1,t}p_{31} - u_{2,t}p_{32} + \epsilon_{3,t}, \\ &\vdots \end{aligned}$$

Stacking the dependent variables into $\breve{\boldsymbol{u}} = (\boldsymbol{u}_2', \ldots, \boldsymbol{u}_m')'$ and the explanatory variables into $\boldsymbol{W} = \operatorname{diag}(-\boldsymbol{u}_1, [-\boldsymbol{u}_1, -\boldsymbol{u}_2], \ldots, [-\boldsymbol{u}_1, \ldots, -\boldsymbol{u}_{m-1}])$ we can write the system as

$$\breve{\boldsymbol{u}} = \boldsymbol{W}\boldsymbol{p} + \breve{\epsilon}$$

where $\breve{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\epsilon)$. Standard calculations then yield the full conditional posterior for $\boldsymbol{p}$ as

$$\boldsymbol{p}|\boldsymbol{\beta}, \boldsymbol{\Psi}_t, \boldsymbol{\xi}, \boldsymbol{h}, \boldsymbol{Y} \sim \mathcal{N}(\overline{\boldsymbol{p}}, \overline{\boldsymbol{V}}_p) \tag{23}$$

for

$$\begin{aligned} \overline{\boldsymbol{V}}_p^{-1} &= \underline{\boldsymbol{V}}_p^{-1} + \boldsymbol{W}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{W}, \\ \overline{\boldsymbol{p}} &= \overline{\boldsymbol{V}}_p \left[\underline{\boldsymbol{V}}_p^{-1}\underline{\boldsymbol{p}} + \boldsymbol{W}'\boldsymbol{\Sigma}_\epsilon^{-1}\breve{\boldsymbol{u}}\right]. \end{aligned}$$

In the MCMC algorithm we replace the draws of $\boldsymbol{\Psi}$ and $\boldsymbol{\Gamma}$ in Algorithm 1 with drawing $\boldsymbol{\beta}$ from (22), the stochastic volatilites and volatility parameters using the `stochvol` package and $\boldsymbol{p}$ from (23).

5 Marginal likelihood and model selection

For the MSM-VAR of Section 3 with constant variance the marginal likelihood is given by

$$\begin{aligned}
m(\boldsymbol{Y}) &= \int p(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\xi})\pi(\boldsymbol{\Gamma}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi})f(\boldsymbol{\xi}|\boldsymbol{w},\tau)\pi(\boldsymbol{w})\pi(\tau)d\boldsymbol{\Gamma}d\boldsymbol{\Psi}d\boldsymbol{\xi}d\boldsymbol{w}d\tau \\
&= \int\left[\int p(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\xi})f(\boldsymbol{\xi}|\boldsymbol{w},\tau)d\boldsymbol{\xi}\right]\pi(\boldsymbol{\Gamma}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi})\pi(\boldsymbol{w})\pi(\tau)d\boldsymbol{\Gamma}d\boldsymbol{\Psi}d\boldsymbol{w}d\tau \\
&= \int p(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{w},\tau)\pi(\boldsymbol{\Gamma}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi})\pi(\boldsymbol{w})\pi(\tau)d\boldsymbol{\Gamma}d\boldsymbol{\Psi}d\boldsymbol{w}d\tau
\end{aligned}$$

where we start by analytically integrating out the latent mixing variable ξ_t. Integrating out ξ_t yields

$$\begin{aligned}
p(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{w},\tau) &= \int p(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\xi})f(\boldsymbol{\xi}|\boldsymbol{w},\tau)d\boldsymbol{\xi} \\
&= \int\left[\prod_{t=1}^{T} p(\boldsymbol{y}_t|\boldsymbol{\Gamma},\boldsymbol{\Psi},\xi_t)f(\xi_t|\boldsymbol{w},\tau)\right]d\xi_1,\ldots,d\xi_T \\
&= \prod_{t=1}^{T}\int p(\boldsymbol{y}_t|\boldsymbol{\Gamma},\boldsymbol{\Psi},\xi_t)f(\xi_t|\boldsymbol{w},\tau)d\xi_t \\
&= \prod_{t=1}^{T}\sum_{i=1}^{K} w_i p_i(\boldsymbol{y}_t|\boldsymbol{\Gamma},\boldsymbol{\Psi},\tau)
\end{aligned}$$

where $p_i(\boldsymbol{y}_t|\boldsymbol{\Gamma},\boldsymbol{\Psi},\tau)$ represents the normal, t and Laplace densities. It then remains to evaluate the relatively low dimensional integral

$$m(\boldsymbol{Y}) = \int p(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{w},\tau)\pi(\boldsymbol{\Gamma}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi})\pi(\boldsymbol{w})\pi(\tau)d\boldsymbol{\Gamma}d\boldsymbol{\Psi}d\boldsymbol{w}d\tau$$

numerically. This can, for example, be done using the method of Chib and Jeliazkov (2001) or using the modifiled harmonic mean estimator (Gelfand and Dey, 1994; Geweke, 1999). See Appendix A for details.

5.1 Stochastic Volatility

In principle the Savage-Dickey density ratio could be used to estimate the Bayes factor in favor of homoskedasticity against stochastic volatility. A homoskedastic version of the VAR with stochastic volatility is obtained by setting $\breve{\sigma}_i^2$ to zero. The basic Savage-Dickey density ratio (Dickey, 1971) states that (under certain conditions) the Bayes factor can be obtained as

$$\mathcal{B}_{H\text{ vs. }SV} = \frac{m_H(\boldsymbol{Y})}{m_{SV}(\boldsymbol{Y})} = \frac{p_{SV}(\breve{\sigma}_i^2 = 0, i = 1\ldots,m)}{\pi_{SV}(\breve{\sigma}_i^2 = 0, i = 1\ldots,m)} \tag{24}$$

where m denotes the marginal likelihoods of the two models and $p_{SV}(\breve{\sigma}_i^2 = 0, i = 1\ldots,m)$ and $\pi_{SV}(\breve{\sigma}_i^2 = 0, i = 1\ldots,m)$ is the marginal posterior and prior for the model with stochastic volatility evaluated at $\breve{\sigma}_i^2 = 0, i = 1\ldots,m$.

In our setting we have the complication that ϕ_i and μ_i are unidentified when $\breve{\sigma}_i^2 = 0$. Koop and Potter (1999) show that (24) holds in the case of unidentified parameters if $\pi_{SV}(\boldsymbol{\theta}|\breve{\sigma}_i^2 = 0, \phi_i^*, \mu_i^*, i = 1\ldots, m) = \pi_H(\boldsymbol{\theta})$ for some value of ϕ_i^* and μ_i^* and $\boldsymbol{\theta}$ the parameters of the homoskedastic model (which are common with the stochastic volatility model). This holds trivially here due to the prior independence between the volatility parameters and the other parameters of the model.

An estimator of the Bayes factor can thus be constructed based on (24). We refrain from doing this. Instead we are content with noting that the shape of the marginal posterior for $\breve{\sigma}_i^2$, in particular the posterior mass at or close to zero, is a valid indicator of the support for stochastic volatility in the data and make informal judgements based on this.

6 Simulation study

In this section, we present the results of estimating the mixture VAR on simulated data with known properties. The data generating process is given by

$$\boldsymbol{y}_t' = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{c}' + \boldsymbol{u}_t',$$

where $\boldsymbol{A}_1 = \text{diag}(0.8)$ and $\boldsymbol{c}$ is taken to be a vector of ones. The error terms $\boldsymbol{u}_t$ can follow multivariate normal, t with 3 degrees of freedom, Laplace distributions or a mixture of these distributions with weights $\boldsymbol{w} = (1/3, 1/3, 1/3)$, with or without stochastic volatility. Several dimensions $m \in \{3, 5, 10\}$ and number of observations $T \in \{100, 200, 400, 1000\}$ are considered.

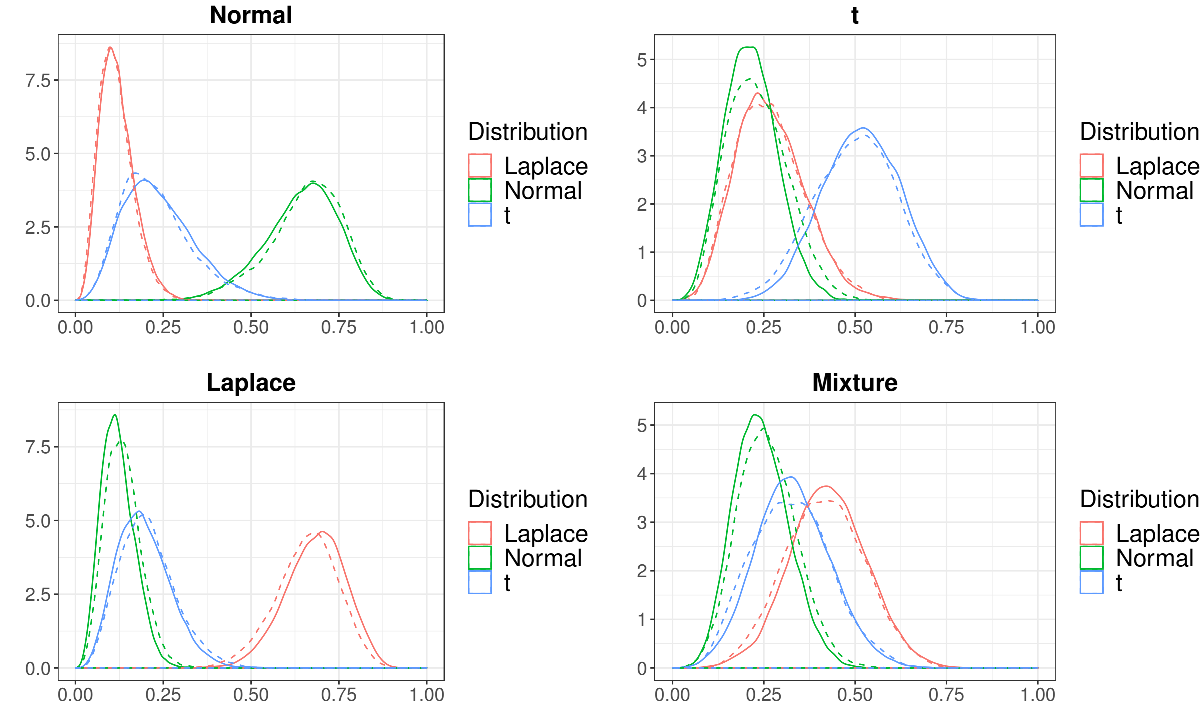
For the MSM-VAR assuming constant variance we use a Minnesota style independent normal Wishart prior (4) where we set the prior mean of $\boldsymbol{\Gamma}$ to zero except for the first own lags where we set the prior mean to 1. The prior standard deviations are set to

$$sd(\gamma_{i,j}) = \begin{cases} \dfrac{\pi_1}{l^{\pi_3}} & \text{for own lags, lag } l \\ \dfrac{\pi_1 \pi_2}{l^{\pi_3}} \dfrac{s_j}{s_k} & \text{lag } l \text{ of variable } k \text{ in equation } j \\ \pi_1 \pi_4 s_j & \text{for the constant term in equation } j \end{cases}$$

with $\pi_1 = 0.2$, $\pi_2 = 0.2$, $\pi_3 =1$, $\pi_4 = 10$ and s_j the residual standard deviation from the OLS estimation of a univariate AR-model for variable j with the same lag length as the estimated VAR. For the inverse Wishart prior for $\boldsymbol{\Psi}$ we set the prior degrees of freedom to $\underline{\nu} = m + 4$ and the scale matrix to $\underline{\boldsymbol{S}} = (\underline{\nu} - m - 1)\,\text{diag}(\boldsymbol{s}^2)$ where m is the number of variables and $\boldsymbol{s}^2$ the vector of residual variances. For the mixture distribution we use the Dirichlet prior $\boldsymbol{w} \sim \mathcal{D}(\underline{\boldsymbol{\alpha}})$ with $\underline{\alpha}_i = 5$ for the weights, slightly favouring equal weights in the mixture. The degrees of freedom, τ, of the t component is endowed with a uniform prior on $(3, 10)$.

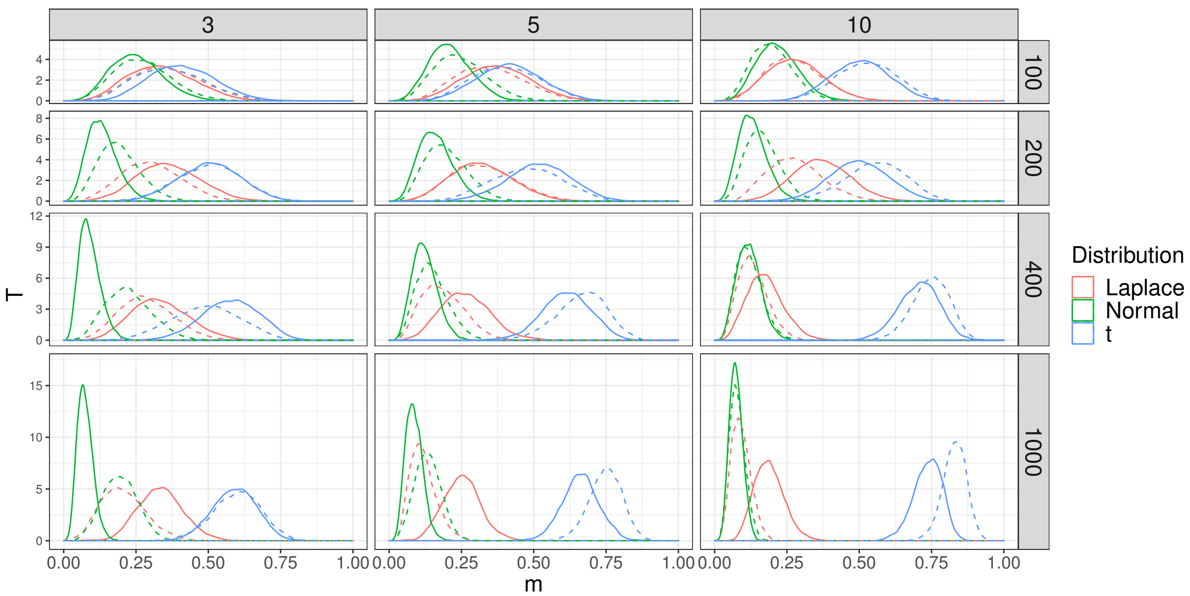
For the MSM-SV-VAR with stochastic volatility we follow the setup in Section 4 with the priors (18), (19) and (20) for the volatility parameters. We set the prior parameters to $(\mu_\mu, \sigma_\mu^2) = (0, 100)$ for an uninformative prior on the constant in (17), $(\alpha_\phi, \beta_\phi) = (5, 1.5)$ for the AR coefficient resulting in prior means $E(\phi_i) = 0.54$ and variance $V(\phi_i) = 0.095$ and $C_{\breve{\sigma}} = 1$, that is a half-normal distribution for $\breve{\sigma}_i$ with variance 1. For the vector $\boldsymbol{p}$ determining the correlation structure we use the normal prior (21) with $\underline{\boldsymbol{p}} = 0$ and $\underline{\boldsymbol{V}}_p = \boldsymbol{I}$.

Figure 2 Posterior distributions of the weights, simulated data, $m = 5$ and $T = 200$



The true distributions of the error terms are generated from a multivariate standard normal, t distribution with 3 degrees of freedom, Laplace and their mixture with weights $\boldsymbol{w} = (1/3, 1/3, 1/3)$. All with constant variance. The posterior distributions of the weights for the MSM-VAR are plotted as solid lines, and with dashed lines for the MSM-SV-VAR.

Figure 3 Posterior distributions of the weights, t-distributed data with 3 degrees of freedom and stochastic volatility



The true distribution of the error terms is generated from a multivariate t distribution with 3 degrees of freedom and stochastic volatility. The posterior distributions of the weights for the MSM-VAR are plotted as solid lines, and with dashed lines for the MSM-SV-VAR. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Table 1 Log marginal likelihood for models without stochastic volatility, simulated data

	Model						
Data	Normal t Laplace	Normal t	Normal Laplace	t Laplace	Normal	t	Laplace
Normal	-1456.5	-1448.8	-1449.7	-1461.8	**-1446.9**	-1458.5	-1496.1
t	-1811.3	-1809.4	-1860.8	-1809.1	-2111.5	**-1807.7**	-1857.5
Laplace	-1362.8	-1386.5	-1357.3	-1354.8	-1498.7	-1385.6	**-1353.1**
Mixture	-1380.8	-1385.9	-1381.1	**-1380.3**	-1468.8	-1386.2	-1382.6

The true distributions of the error terms are generated from homoskedastic multivariate standard normal, t distribution with 3 degrees of freedom, Laplace distributions and their mixture with weights $\boldsymbol{w} = (1/3, 1/3, 1/3)$. The dimension of the VAR is $m = 3$ and the number of observations $T = 200$. The method of Chib and Jeliazkov (2001) in Appendix A is used to estimate the marginal likelihoods.

When estimating the models we include two lags of $\boldsymbol{y}_t$ and take 20,000 draws from the MCMC algorithm after an initial 2,000 draws as burn-in.

Figure 2 shows the posterior distribution of the mixture weights for different true distributions (normal, t, Laplace and the equal weight mixture) of the error terms with $m = 5$ and $T = 200$. In each case we get a clear indication of what the true distribution is with the posterior mean of the weight of the correct component being considerably larger than the other components. For the equal mixture the posterior distributions and the posterior means are quite close together. As can be expected, the results becomes more clear-cut as the amount of information, sample size or dimension of the model, increases. This is illustrated in Figure 3 for t-distributed errors with stochastic volatility[3]. The results for the other error distributions we consider are similar except for the case of normally distributed data with stochastic volatility (see Figures 16-22 in Appendix B). In this case, the model not allowing for stochastic volatility tends to allocate more weight to the t-component of the mixture than the normal component.

More formal model comparisons can be based on marginal likelihoods. Table 1 reports the log marginal likelihoods for different specifications of the mixture in the MSM-VAR model estimated on normal, t, Laplace and mixture distributed data with constant variance. In all cases except for the equal weights mixture the correct specification receives the highest log marginal likelihood. For the equal weights mixture the model with a mixture of t and Laplace distributions receives a marginally higher log marginal likelihood than the true three component mixture.

As alluded to, the posterior distribution of the innovations for the log volatilities can be informative about the presence of stochastic volatility in the data. To illustrate this Figure 4 shows the posterior distribution of the volatility parameters ϕ_1 and $\breve{\sigma}_1^2$ of the MSM-SV-VAR for a case without stochastic volatility (see Figure 23 in Appendix B for the posterior distributions of all volatility parameters). Here we can observe that there is substantial posterior mass at or close to zero in the distribution of the innovation variances. In addition, the AR parameters, ϕ_i, are poorly identified by the data and the posterior distributions are quite close to the prior. The same behavior can be observed when the true distribution of the error terms is generated from a multivariate t with 3 degrees of freedom, Laplace and the equal weight mixture distribution. Additionally, in

[3] Stochastic volatility is simulated via the `R` package `stochvol` with $\mu_i = -10$, $\phi_i = 0.98$ and $\breve{\sigma}_i^2 = 0.2$.

Figure 5, we present the posterior distribution of the volatility parameters ϕ_1 and $\breve{\sigma}_1^2$ for a case of normally distributed data with stochastic volatility (see Figure 24 in Appendix B for the posterior distributions of all volatility parameters). In this case, we get a clear indication of stochastic volatility in the data. Similar behavior is observed when the true distribution of the error terms is generated from a multivariate t with 3 degrees of freedom, Laplace and mixture distributions with stochastic volatility.

Figure 4 Posterior distributions of volatility parameters ϕ_1 and $\breve{\sigma}_1^2$, normally distributed data, $m = 5$ and $T = 200$

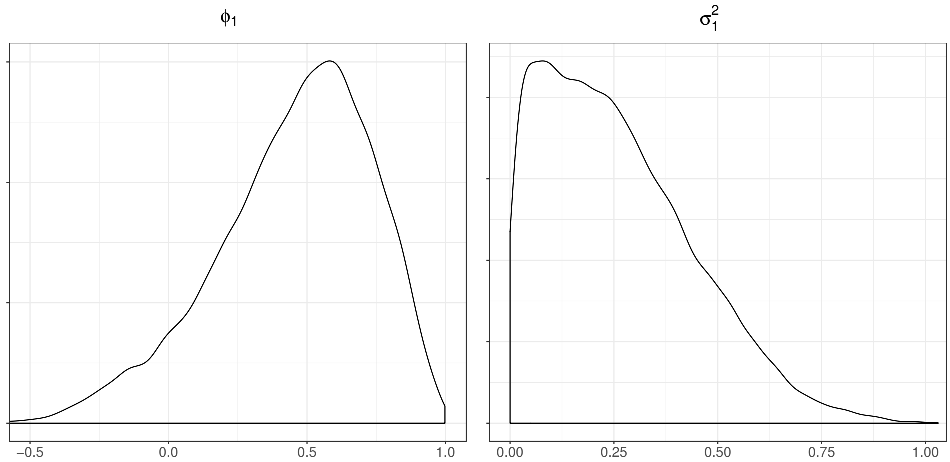
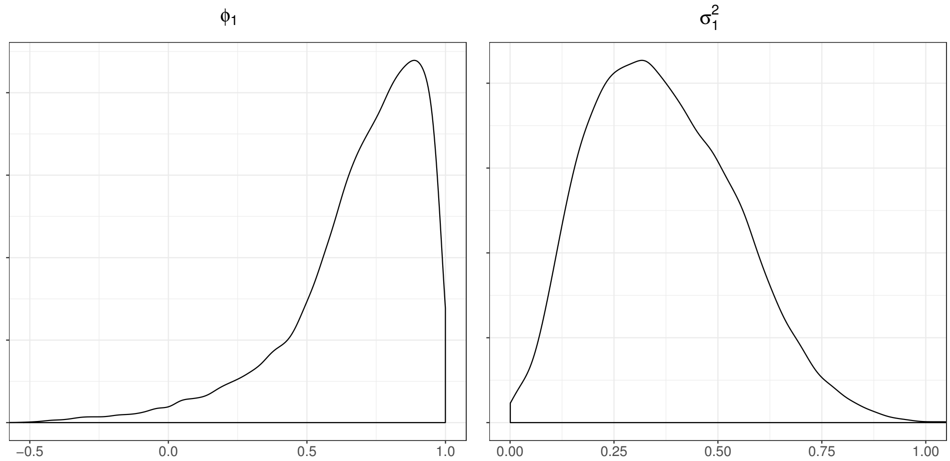


Figure 5 Posterior distributions of volatility parameters ϕ_1 and $\breve{\sigma}_1^2$, normally distributed data with stochastic volatility, $m = 5$ and $T = 200$



7 Empirical results

In this section we apply the methods developed in the paper to the kind of small and medium sized VARs frequently used in applied macroeconomics. The first, 3 variable VAR, is inspired by Cogley and Sargent (2005) and Primiceri (2005) and the second, 7 variable VAR is a variation on the VAR of Smets and Wouters (2007).

Table 2 Log marginal likelihood for 3-variable VAR without stochastic volatility

Country	Model						
	Normal t Laplace	Normal t	Normal Laplace	t Laplace	Normal	t	Laplace
Austria	441.8	442.1	434.7	443.0	371.2	**444.8**	433.3
Canada	-676.0	-675.2	-697.3	**-671.4**	-894.0	**-671.1**	-700.9
Denmark	100.7	102.3	38.8	102.1	-165.6	**105.5**	40.4
Finland	240.0	243.9	195.8	249.1	-78.9	**255.3**	207.1
France	547.1	546.4	530.5	548.0	404.6	**549.6**	527.9
Germany	623.2	621.1	620.3	**623.9**	536.1	622.9	619.5
Italy	153.7	152.7	140.2	**154.3**	-5.4	**154.2**	139.1
Japan	624.0	624.8	614.8	626.4	548.4	**628.9**	616.5
South Korea	-303.8	-303.6	-322.8	-298.1	-485.2	**-295.7**	-315.4
Netherlands	620.7	618.8	617.5	**621.5**	554.5	620.5	609.6
Norway	0.4	2.8	-47.2	1.9	-251.3	**6.3**	-46.1
Spain	55.8	54.5	42.1	**60.6**	-95.5	**60.8**	48.0
Sweden	-406.4	-403.3	-465.5	-401.2	-790.9	**-395.5**	-458.9
United Kingdom	423.8	423.7	413.7	427.6	300.1	**429.2**	417.1
United States	-218.3	-213.1	-267.6	-212.8	-500.9	**-205.3**	-268.5

All models estimated using 50 000 draws from the sampler and 5 000 draws as burn-in. The method of Chib and Jeliazkov (2001) in Appendix A is used to estimate the marginal likelihoods.

7.1 A small VAR

We estimate a small VAR similar to Cogley and Sargent (2005) and Primiceri (2005) for the 15 OECD countries in Table 2 using monthly data. The variables are inflation, unemployment rate and a 3-month interest rate sourced from the OECD Main Economic Indicators. Details on the variables and sample periods for the different countries are given in Appendix C. We use a lag length of 6 for all countries and the prior set up is the same as in Section 6 with the exception that we set the prior means of the first own lags in $\boldsymbol{\Gamma}$ to 0.8, reflecting a belief that the data is stationary.

Table 2 reports the log marginal likelihoods for the 3-component MSM-VAR with constant variance as well as the special cases that can be obtained by turning off different components in the mixture. The table gives a very clear message that the normality assumption is not supported by the data. While our 3 component mixture with normal, t and Laplace components is never the preferred specification it is clearly superior to a pure normal likelihood. While a pure t likelihood is the preferred specifications for most of the countries, this not the case for all of them. A mixture of t and Laplace distributions is preferred for Germany and the Netherlands and there is a tie between the t-Laplace mixture and a pure t likelihood for Canada, Spain and Italy – something which would have been difficult to detect without our flexible modeling framework.

To provide additional insights we take a closer look at the German data displayed in Figure 6. We start by viewing the data through the lens of the constant variance 3-component MSM-VAR. The posterior distribution of the weights of the normal, t, and Laplace components in the mixture is shown in Figure 7(a). There is strong and equal support for the t and Laplace components with the posterior distribution of the weights

Figure 6 Data for the small VAR - Germany

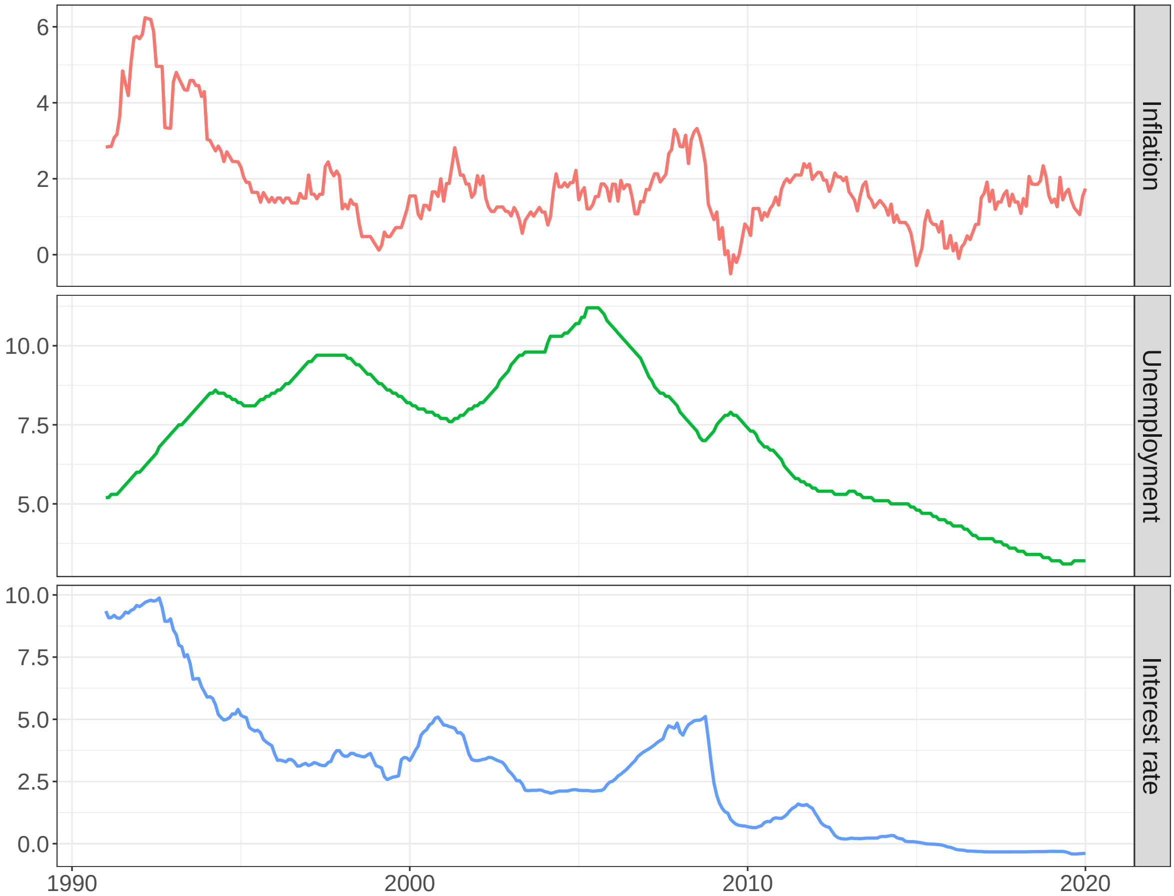
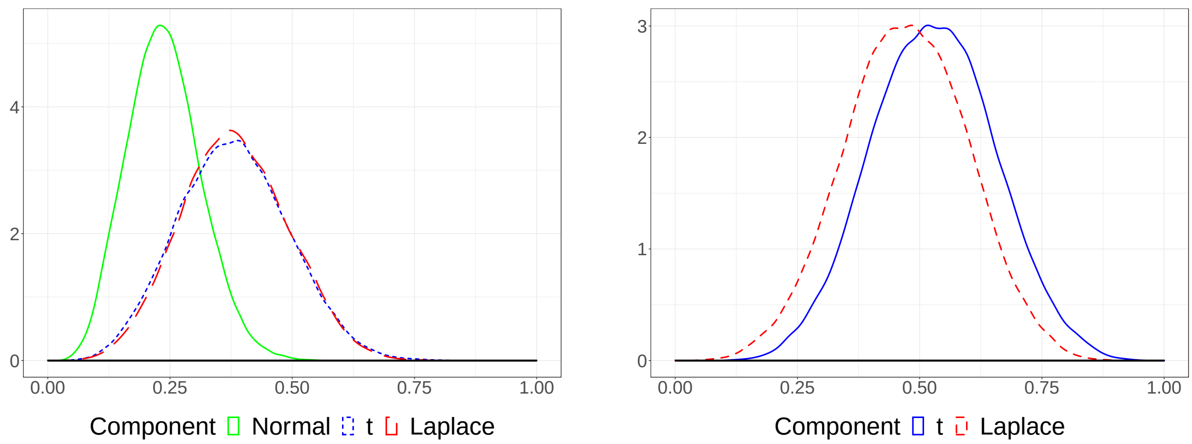


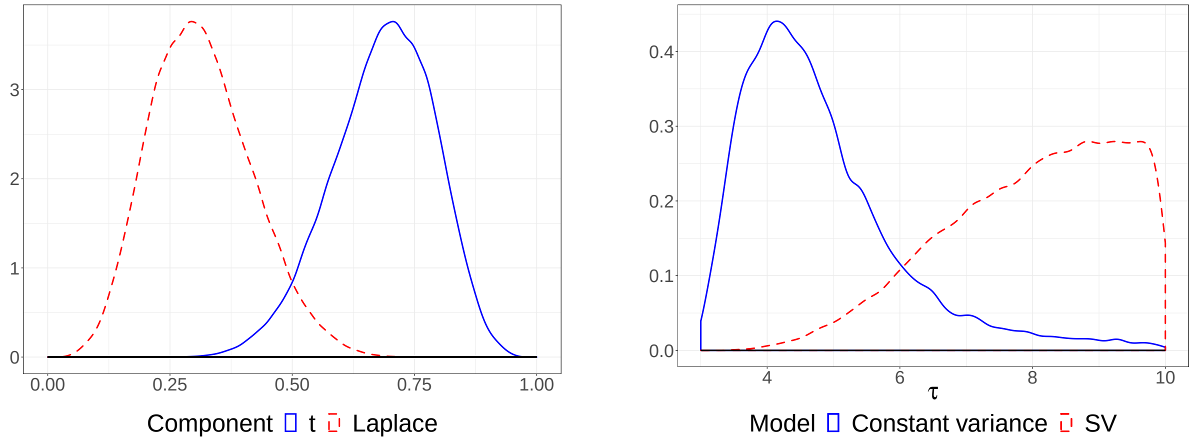
Figure 7 Posterior distribution of mixture weights, small VAR - Germany



(a) Three component mixture

(b) Two component mixture

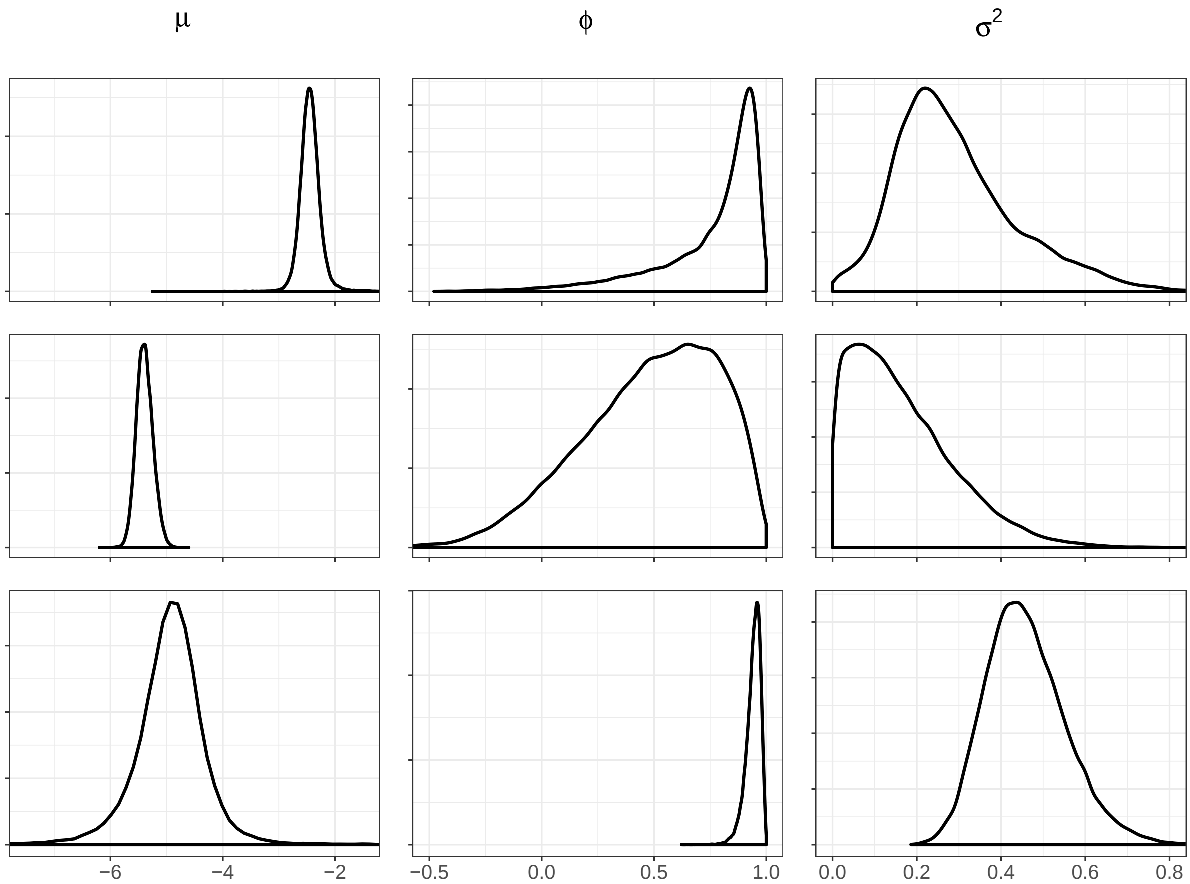
Figure 8 Posterior distribution of mixture parameters, small VAR - Germany



(a) mixture weigths, Stochastic volatility

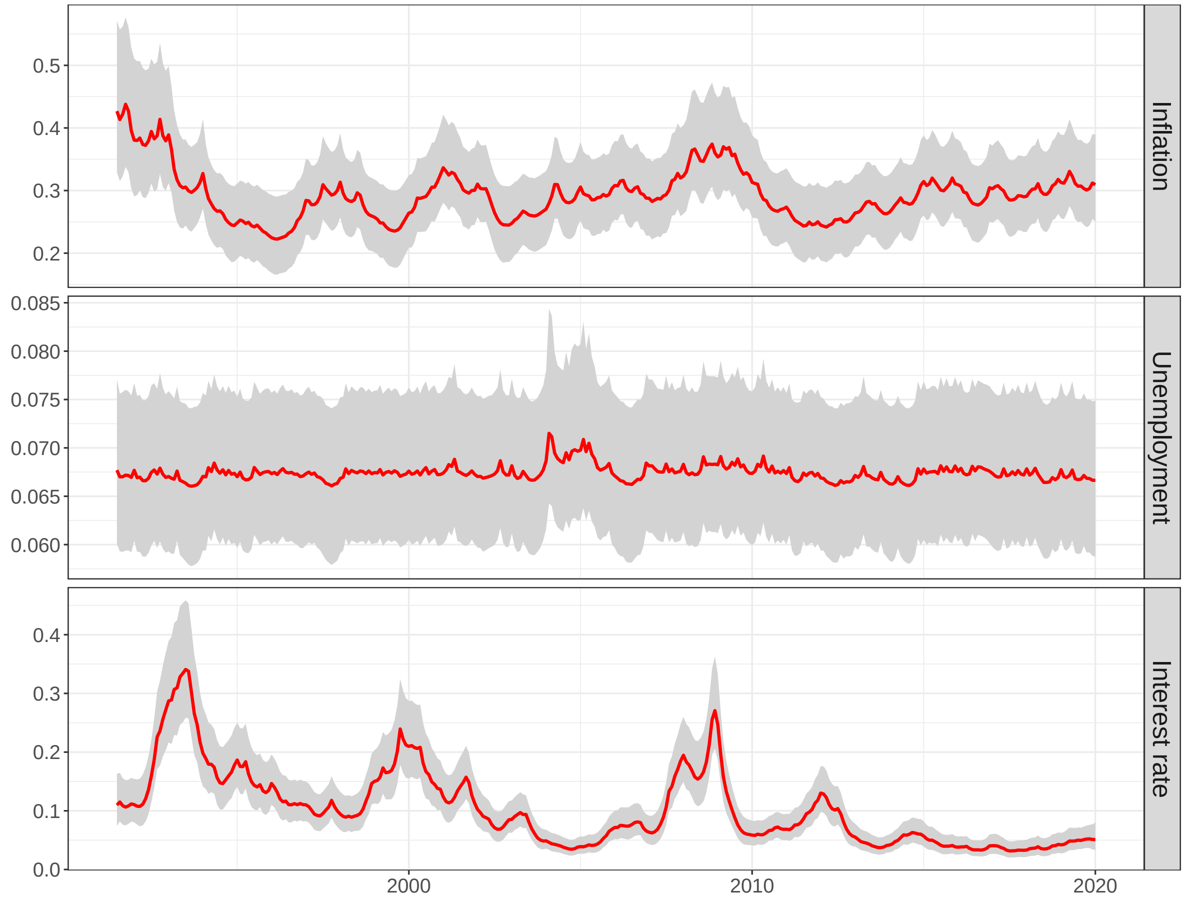
(b) τ, constant variance and stochastic volatility

Figure 9 Posterior distribution of volatility parameters, small VAR - Germany



Posterior distribution of μ_i, ϕ_i and $\breve{\sigma}_i^2$ in the columns. Row 1 is equation for Inflation, row 2 Unemployment and row 3 the interest rate.

Figure 10 Posterior distribution of volatilities, small VAR - Germany



Solild line is median volatility (standard deviation) and shaded area 68% credible region.

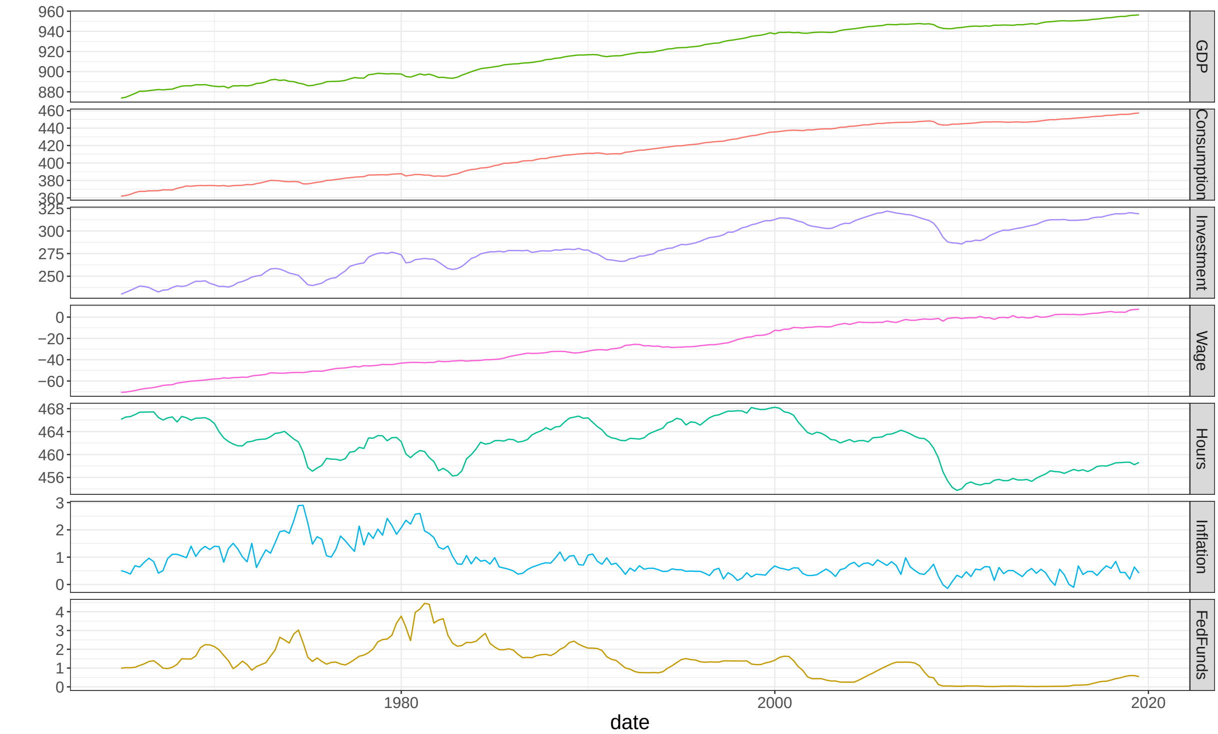
being almost identical and less support for the normal component. As we know from Table 2 the t-Laplace mixture is the preferred specification and re-estimating the model with only a t and Laplace component in the mixture again yields essentially equal weights for the t and Laplace components, see Figure 7(b).

Next we investigate the possibility that the fat tails captured by the t and Laplace components are, in fact, due to time-varying volatilities. Turning to the MSM-SV-VAR of Section 4 we use the same prior set up for the volatility part as in Section 6.

Figure 8(a) shows the posterior distribution of the mixture weights when allowing for stochastic volatility. The weight of the t component has increased compared to the model without stochastic volatility and the posterior mean has moved from 0.53 to 0.69. The posterior distribution of τ, the degrees of freedom for the t distribution is displayed in Figure 8(b) for the model with and without stochastic volatility. Allowing for stochastic volatility leads to an increase in the degrees of freedom and the mixture distribution is now less fat-tailed as the stochastic volatility accounts for some – but not all – of the non-normality of the distribution.

Turning to the volatilities, the posterior distribution of the volatility parameters is shown in Figure 9. The posterior distribution of the variance of the volatility innovations, $\breve{\sigma}_i^2$, shows a substantial posterior mass at zero for unemployment indicating that a constant variance is preferred for this equation. The posterior distribution of the variance of the

Figure 11 Data for the medium VAR



volatility innovations for the interest rate is, on the other hand, well separated from zero showing strong support for time varying volatility in this equation. For inflation we have a middle ground with some posterior mass at zero but overall support for time varying volatility. Turning to the autoregressive parameter, ϕ_i, we note a near unit root behaviour for the log volatilities of the interest rate equation whereas they appear to be stationary for inflation and for unemployment the posterior distribution – reflecting the lack of time varying volatility – mimics the prior. The posterior distribution of the volatilities is displayed in Figure 10. For the inflation equation the volatilities are relatively stable over time with peaks at the beginning of the sample immediately after German unification and around the financial crisis. The volatility of unemployment is, as expected, stable over time. For the interest rate the volatility is quite small for most of the period with a large peak shortly after German unification and smaller peaks around the year 2000 and at the financial crisis. The very small volatility at the end of the sample is clearly influenced by the low interest rate environment.

Overall we find support for both non-normality and stochastic volatility and that the non-normality can be captured by a mixture of t and Laplace distributions.

7.2 A medium sized VAR

Smets and Wouters (2007) estimated a VAR with seven variables for the USA, real GDP, real consumption, real investments, the real wage, hours worked, inflation and the federal funds rate on data from 1966Q1 to 2004Q4. We extend the data set up to 2019Q3, see Appendix C for data sources and transformations, and display the variables in Figure 11. We use the same prior setup as in the previous section with the exception that we set the prior means of the first own lag to one for the first four variables and the remaining to

Figure 12 Posterior distribution of mixture weights, medium VAR

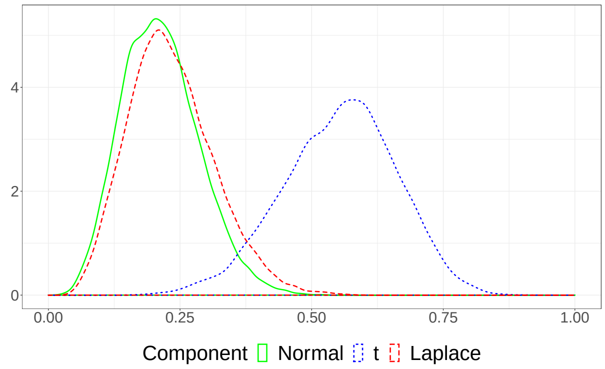


Table 3 Log marginal likelihood for models without stochastic volatility - Medium sized VAR

Model						
Normal t Laplace	Normal t	Normal Laplace	t Laplace	Normal	t	Laplace
-1113.4	-1103.0	-1192.1	-1112.4	-1206.3	**-1099.0**	-1319.0

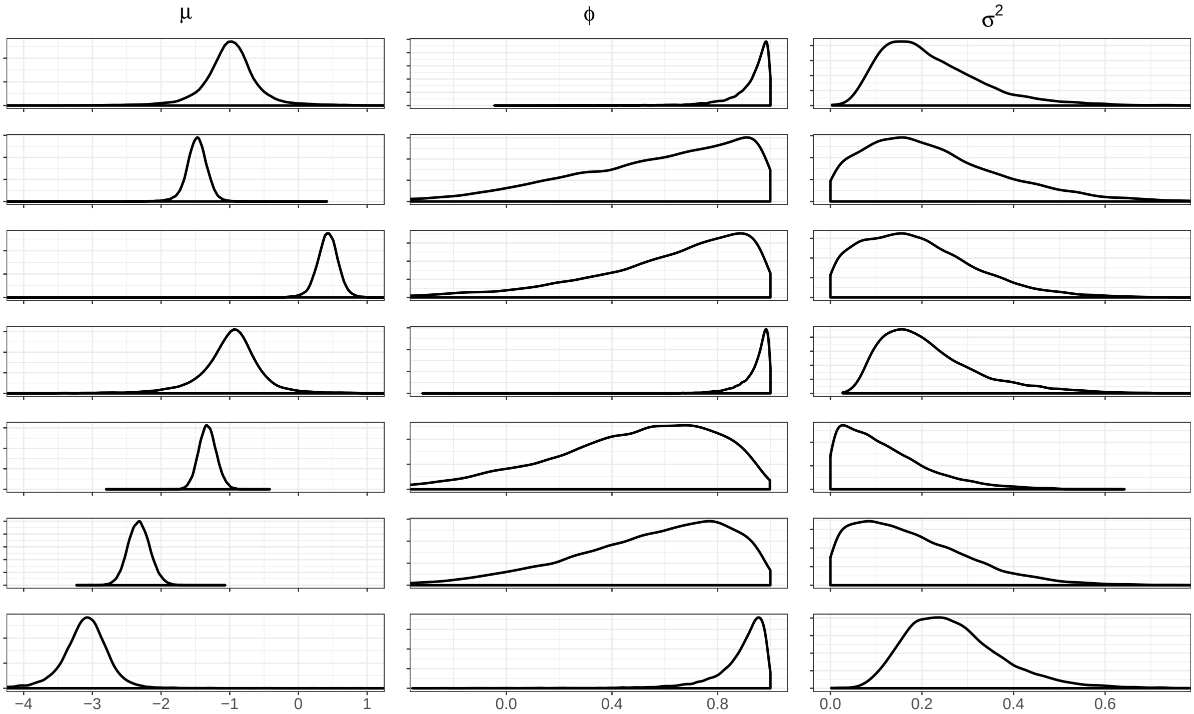
0.8 and start by estimating the MSM-VAR with constant variance.

The posterior distribution of the weights in the mixture is shown in Figure 12. We note that the t component dominates with less support for the normal and Laplace components. The posterior mean of the weights are 0.21, 0.55 and 0.23 for the normal, t and Laplace components. This picture is confirmed by the marginal likelihoods in Table 3 where the model with only a t distribution is preferred.

Turning to the issue of time-varying volatilities we continue with the pure t distribution and add stochastic volatility. The posterior distributions of the volatility parameters are displayed in Figure 13. The posterior distributions for $\breve{\sigma}_i^2$ indicate limited support in the data for stochastic volatility as all the posterior distributions except for the GDP, wage and federal funds rate equations have substantial posterior mass at or close to zero. Corresponding to this we see that there is little information in the data about ϕ_i and the posterior is close to the prior except for the GDP, wage and federal funds rate equation. A similar picture is painted by the estimated volatilities in Figure 14 which show very little time variation for consumption, investments, hours worked and inflation. Figure 15, finally, shows the posterior distribution of the the degrees of freedom, τ, for the model with and without stochastic volatility. For this data set there is small shift to the right (the posterior mean increases from 5.1 to 6.3) and thinner tails when we add stochastic volatility, again indicating that the contribution of the stochastic volatility is small in terms of capturing the fat tails.

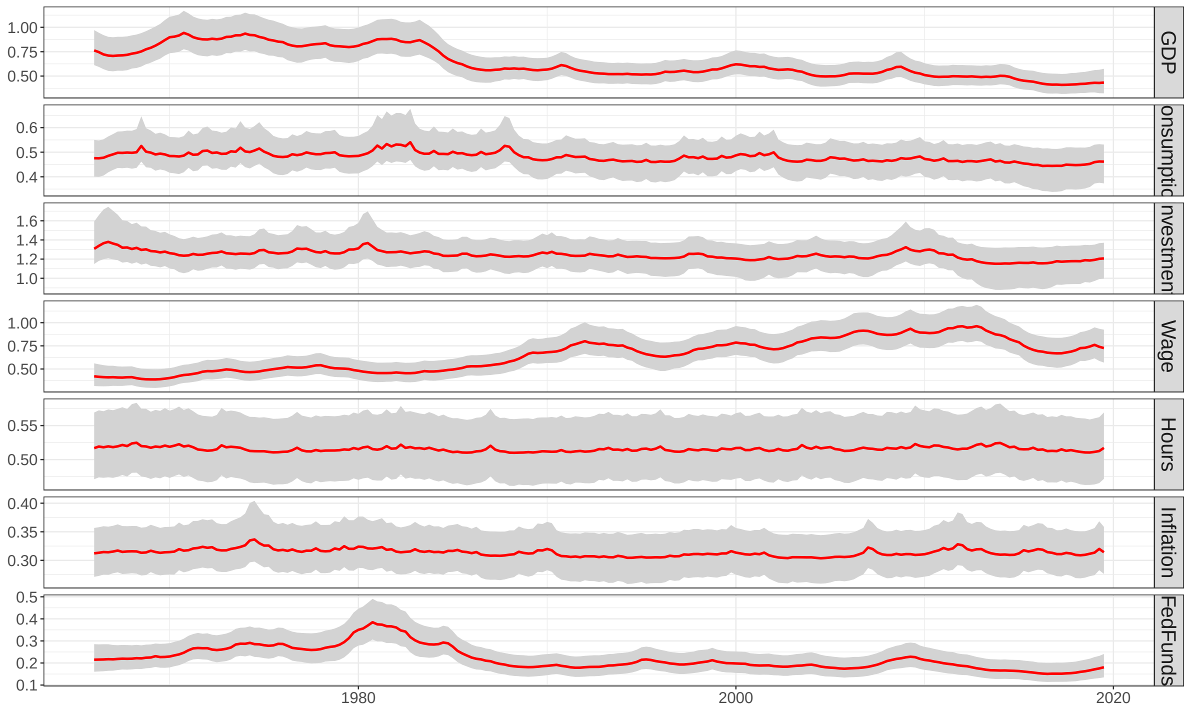
We thus find support for a non-normal error distribution and the set up with a mixture distribution has allowed us to identify the t distribution as a plausible candidate. In addition, while there is support for stochastic volatility in some of the equations, the contribution of the stochastic volatility to the fatness of the tails seems to be limited in this case.

Figure 13 Posterior distribution of volatility parameters, medium VAR



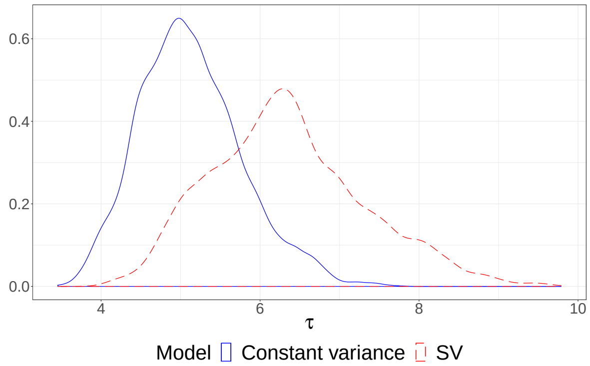
Columns show the volatility parameters, μ_i, ϕ_i and $\breve{\sigma}_i^2$. Rows corresponds to equations, real GDP, real consumption, real investments, real wage, hours worked, inflation and the federal funds rate.

Figure 14 Posterior distribution of volatilities, medium VAR



Solid line is median volatility (standard deviation) and shaded area 68% credible region.

Figure 15 Posterior distribution for τ, constant variance and stochastic volatility, medium VAR



8 Conclusions

This paper suggests a simple and flexible framework for modeling non-normal data. This is done by considering a discrete mixture of scale mixtures of normal distributions leading to the MSM-VAR. The proposed framework can adapt to a wide range of distributional shapes while containing the normal, t and Laplace distribution as special cases. The use of scale mixtures of normals leads to feasible and computationally straightforward MCMC implementations for Bayesian inference. In addition, the frameworks is easily extended to include additional or alternative distributions that can be represented as a scale mixture of a normal distribution.

Since fat tails in the distribution of the error terms can be difficult to distinguish from time-varying volatility, we extended all the results to the Bayesian VAR-model which allows for stochastic volatility in the error distribution (MSM-SV-VAR). Using simulated data and in applications to macroeconomic data we show how marginal likelihoods and other features of the posterior distribution can be used for model selection.

In the applications we find support for both fat tails and stochastic volatility in the macroeconomic data. Both are thus important features of the data that should be accounted for in empirical work. In addition we find that the non-normality of the data in many cases is better modeled with a mixture of distributions rather than, as is frequently the case, just assuming a t distribution instead of the normal distribution.

In future research, we plan to extend our results to flexible models that allow for fat tails as well as skewness in the error distribution.

Acknowledgement

The authors acknowledge financial support from the project ”Models for macro and financial economics after the financial crisis” (Dnr: P18-0201) funded by the Jan Wallander and Tom Hedelius Foundation. Stepan Mazur also acknowledges financial support from the internal research grants of Örebro University.

Appendices

A Marginal likelihood for the MSM-VAR

A.1 Chib-Jeliazkov

The method of Chib and Jeliazkov (2001) is based on the basic marginal likelihood identity which gives the marginal likelihood as

$$m(\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*, \tau^*)\pi(\boldsymbol{\Gamma}^*|\boldsymbol{\Psi}^*)\pi(\boldsymbol{\Psi}^*)\pi(\boldsymbol{w}^*)\pi(\tau^*)}{p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*, \tau^*|\boldsymbol{Y})}$$

where the expression holds for any value of $\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*$ and τ^*. The numerator is straightforward to evaluate and the denominator can be estimated with MCMC.

Write $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*, \tau^*|\boldsymbol{Y}) = p(\tau^*|\boldsymbol{Y})p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*\boldsymbol{w}^*|\tau^*, \boldsymbol{Y})$. For $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ and $\boldsymbol{w}$ the full conditional posteriors are available in closed form but not for τ which is sampled with a Metropolis-Hastings step. $p(\tau^*|\boldsymbol{Y})$ can be estimated using the method of Chib and Jeliazkov (2001) using the output from the full MCMC run and a second run with τ fixed at τ^*. Let $\alpha(\tau, \tau^*|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{Y})$ be the acceptance probability (16) in the M-H step proposing a move from τ to τ^* and $q(\tau, \tau^*|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{Y})$ the proposal distribution. We then have

$$p(\tau^*|\boldsymbol{Y}) = \frac{E_1\left[\alpha(\tau, \tau^*|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{Y})q(\tau, \tau^*|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{Y})\right]}{E_2\left[\alpha(\tau^*, \tau|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{Y})\right]}$$

where the expectation E_1 is with respect to the posterior distribution of $(\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \tau, \boldsymbol{\xi}, \boldsymbol{s})$ end E_2 with respect to the distribution $p(\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}|\tau^*, \boldsymbol{Y})q(\tau^*, \tau|\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{Y})$. Draws from the latter is obtained by running the original MCMC sampler with τ fixed at τ^* and generating a proposal τ from $q(\cdot)$ given τ^* for each draw of $(\boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{s})$.

An estimate of $p(\tau^*|\boldsymbol{Y})$ is then given by

$$\widehat{p}(\tau^*|\boldsymbol{Y}) = \frac{\frac{1}{R}\sum_{i=1}^{R} \alpha(\tau^{(i)}, \tau^*|\boldsymbol{\xi}^{(i)}, \boldsymbol{s}^{(i)}, \boldsymbol{Y})q(\tau^{(i)}, \tau^*|\boldsymbol{Y})}{\frac{1}{R}\sum_{j=1}^{R} \alpha(\tau^*, \tau^{(j)}|\boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)}, \boldsymbol{Y})}$$

where we have used that the proposal is a symmetric random walk proposal that does not depend on the other parameters and that the acceptance probability only depends on $\boldsymbol{\xi}$ and $\boldsymbol{s}$.

A.1.1 Normal-Wishart prior

For $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{Y})$ we note that the draws of $(\boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)})$ from the reduced run with $\tau = \tau^*$ has the marginal distribution $p(\boldsymbol{\xi}, \boldsymbol{s}|\tau^*, \boldsymbol{Y})$ and can be used to estimate $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\tau^*, \boldsymbol{Y})$ as

$$\widehat{p}(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\tau^*, \boldsymbol{Y}) = \frac{1}{R}\sum_{j=1}^{R} p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\tau^*, \boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)}, \boldsymbol{Y}).$$

Normally $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{Y})$ would have to be estimated in two steps when they are sampled in two blocks. The conditional structure offer a simplification here and we can estimate $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{Y})$ in one step since

$$\begin{aligned} p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\tau^*, \boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)}, \boldsymbol{Y}) &= p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*|\boldsymbol{w}^*, \tau^*, \boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)}, \boldsymbol{Y})p(\boldsymbol{w}^*|\tau^*, \boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)}, \boldsymbol{Y}) \\ &= p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*|\boldsymbol{\xi}^{(j)}, \boldsymbol{Y})p(\boldsymbol{w}^*|\boldsymbol{s}^{(j)}, \boldsymbol{Y}). \end{aligned}$$

A.1.2 Independent normal Wishart prior

With the independent normal Wishart prior $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*|\boldsymbol{w}^*, \tau^*, \boldsymbol{\xi}^{(j)}, \boldsymbol{s}^{(j)}, \boldsymbol{Y})$ is not available in closed form and additional reduced MCMC rounds are needed. Write

$$p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\tau^*, \boldsymbol{Y}) = p(\boldsymbol{w}^*|\tau^*, \boldsymbol{Y})p(\boldsymbol{\Gamma}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{Y})p(\boldsymbol{\Psi}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{\Gamma}^*, \boldsymbol{Y}).$$

We can then estimate $p(\boldsymbol{w}^*|\tau^*, \boldsymbol{Y})$ as

$$\widehat{p}(\boldsymbol{w}^*|\tau^*, \boldsymbol{Y}) = \frac{1}{R}\sum_{j=1}^{R} p(\boldsymbol{w}^*|\tau^*, \boldsymbol{s}^{(j)}, \boldsymbol{Y})$$

using the output from the reduced run with $\tau = \tau^*$. For the second reduced run we also fix $\boldsymbol{w} = \boldsymbol{w}^*$ and use the output to estimate $p(\boldsymbol{\Gamma}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{Y})$ as

$$\widehat{p}(\boldsymbol{\Gamma}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{Y}) = \frac{1}{R}\sum_{j=1}^{R} p(\boldsymbol{\Gamma}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{\Psi}^{(j)}, \boldsymbol{\xi}^{(j)}, \boldsymbol{Y}).$$

A final reduced MCMC run additionally fixes $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^*$ and the output is used to estimate $p(\boldsymbol{\Psi}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{\Gamma}^*, \boldsymbol{Y})$ as

$$\widehat{p}(\boldsymbol{\Psi}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{\Gamma}^*, \boldsymbol{Y}) = \frac{1}{R}\sum_{j=1}^{R} p(\boldsymbol{\Psi}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{\Gamma}^*, \boldsymbol{\xi}^{(j)}, \boldsymbol{Y}).$$

A.1.3 When there is no t component in the mixture (or τ is not estimated)

The problem simplifies. We are then concerned with

$$m(\boldsymbol{Y}) = \frac{m(\boldsymbol{Y}|\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*)\pi(\boldsymbol{\Gamma}^*|\boldsymbol{\Psi}^*)\pi(\boldsymbol{\Psi}^*)\pi(\boldsymbol{w}^*)}{p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{Y})}.$$

The numerator is available in closed form and we are left with estimating $p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{Y})$ which can be done using the output from the original MCMC run as

$$\widehat{p}(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{Y}) = \frac{1}{R}\sum_{i=1}^{R}\sum_{i=1}^{R} p(\boldsymbol{\Gamma}^*, \boldsymbol{\Psi}^*, \boldsymbol{w}^*|\boldsymbol{\xi}^{(i)}, \boldsymbol{s}^{(i)}, \boldsymbol{Y})$$

with the Normal-Wishart prior. For the independent normal Wishart prior the original run is used to estimate $p(\boldsymbol{w}^*|\tau^*, \boldsymbol{Y})$ and two reduced runs are needed for estimating $p(\boldsymbol{\Gamma}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{Y})$ and $p(\boldsymbol{\Psi}^*|\tau^*, \boldsymbol{w}^*, \boldsymbol{\Gamma}^*, \boldsymbol{Y})$.

A.2 Modified harmonic mean

The modified harmonic mean estimator of the marginal likelihood (Gelfand and Dey, 1994) uses the relation

$$\frac{1}{m(\boldsymbol{Y})} = \frac{\int f(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{m(\boldsymbol{Y})} = \int \frac{f(\boldsymbol{\theta})}{m(\boldsymbol{Y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})} p(\boldsymbol{\theta}|\boldsymbol{Y})\, d\boldsymbol{\theta}$$

to estimate the marginal likelihood as

$$\widehat{m}(\boldsymbol{Y}) = \left[\frac{1}{R}\sum_{i=1}^{R} \frac{f\left(\boldsymbol{\theta}^{(i)}\right)}{m\left(\boldsymbol{Y}|\boldsymbol{\theta}^{(i)}\right)\pi\left(\boldsymbol{\theta}^{(i)}\right)}\right]^{-1}$$

using draws $\boldsymbol{\theta}^{(i)} = \left(\boldsymbol{\Gamma}^{(i)}, \boldsymbol{\Psi}^{(i)}, \boldsymbol{w}^{(i)}, \tau^{(i)}\right)$ from the posterior. A necessary condition is that f is a proper density with support contained in the parameter space of the model under consideration.

For an accurate estimate we need a good choice of the density f that approximates $m(\boldsymbol{Y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$ well. Geweke (1999) suggests using a truncated multivariate normal. For this to work well we need to transform $\boldsymbol{\theta}$ (that is, $\boldsymbol{\Psi}$, $\boldsymbol{w}$ and τ) to an unrestricted parameter space.

For $\boldsymbol{\Psi}$ we use the matrix logarithm, $\widetilde{\boldsymbol{\Psi}} = \ln\boldsymbol{\Psi} = \boldsymbol{Q}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{Q}'$, with inverse transformation $\boldsymbol{\Psi} = \exp(\widetilde{\boldsymbol{\Psi}}) = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$ where $\boldsymbol{Q}$ is the matrix of eigenvectors of $\boldsymbol{\Psi}$ and $\widetilde{\boldsymbol{\Psi}}$ and $\boldsymbol{\Lambda}$ is the matrix with the eigenvalues of $\boldsymbol{\Psi}$ on the diagonal and $\widetilde{\boldsymbol{\Lambda}}$ has the logarithm of the eigenvalues on the diagonal. That is, the eigenvalues of $\widetilde{\boldsymbol{\Psi}}$ are the logarithm of eigenvalues of $\boldsymbol{\Psi}$. These expressions are valid for symmetric matrices and since it is sufficient to consider the unique elements of $\boldsymbol{\Psi}$ we focus on the transformation from $\boldsymbol{\psi} = \operatorname{vech}\boldsymbol{\Psi}$ to $\widetilde{\boldsymbol{\psi}} = \operatorname{vech}\widetilde{\boldsymbol{\Psi}}$.

Following Linton and McCrorie (1995) we have

$$\operatorname{dvec}\boldsymbol{\Psi} = (\boldsymbol{Q}\otimes\boldsymbol{Q})\,\boldsymbol{H}\,(\boldsymbol{Q}\otimes\boldsymbol{Q})'\operatorname{dvec}\widetilde{\boldsymbol{\Psi}}$$

with $\boldsymbol{H} = \operatorname{diag}(h_i), i = 1,\ldots,m^2$, and

$$h_i = \begin{cases} \dfrac{e^{\widetilde{\lambda}_k} - e^{\widetilde{\lambda}_j}}{\widetilde{\lambda}_k - \widetilde{\lambda}_j} = \dfrac{\lambda_k - \lambda_j}{\ln\lambda_k - \ln\lambda_j}, & i = (k-1)m + j, k \neq j \\ e^{\widetilde{\lambda}_k} = \lambda_k, & k = j \end{cases}$$

where $\widetilde{\lambda}_k$ and λ_k are the eigenvalues of $\widetilde{\boldsymbol{\Psi}}$ and $\boldsymbol{\Psi}$. The Jacobian matrix for the transformation from $\widetilde{\boldsymbol{\psi}}$ to $\boldsymbol{\psi}$ is then

$$\boldsymbol{J}_{\widetilde{\boldsymbol{\psi}}} = \frac{\partial\boldsymbol{\psi}}{\partial\widetilde{\boldsymbol{\psi}}} = \boldsymbol{D}_m^{+}(\boldsymbol{Q}\otimes\boldsymbol{Q})\,\boldsymbol{H}\,(\boldsymbol{Q}\otimes\boldsymbol{Q})'\boldsymbol{D}_m$$

where the duplication matrix $\boldsymbol{D}_m$ satisfies $\operatorname{vec}\widetilde{\boldsymbol{\Psi}} = \boldsymbol{D}_m\operatorname{vech}\widetilde{\boldsymbol{\Psi}}$ and $\boldsymbol{D}_m^{+} = (\boldsymbol{D}_m'\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m'$ satisfies $\operatorname{vech}\boldsymbol{\Psi} = \boldsymbol{D}_m^{+}\operatorname{vec}\boldsymbol{\Psi}$.

For $\boldsymbol{w}$, we note that it is confined to the $K-1$ simplex and

$$w_i = g_{\boldsymbol{w}}^{-1}(\widetilde{\boldsymbol{w}}) = \begin{cases} \dfrac{e^{\widetilde{w}_i}}{1+\sum_{j=1}^{K-1} e^{\widetilde{w}_j}}, & i = 1,\ldots,K-1 \\ 1 - \displaystyle\sum_{j=1}^{K-1} w_i = \dfrac{1}{1+\sum_{j=1}^{K-1} e^{\widetilde{w}_j}}, & i = K \end{cases}$$

maps $\widetilde{\boldsymbol{w}} \in \mathbb{R}^{K-1}$ to the simplex. The mapping from $\boldsymbol{w}$ to $\widetilde{\boldsymbol{w}}$ is

$$\widetilde{w}_i = g_{\boldsymbol{w}}(\boldsymbol{w}) = \ln(w_i/w_K), \quad i = 1, \ldots, K-1$$

with Jacobian elements

$$\frac{\partial \widetilde{w}_i}{\partial w_j} = \begin{cases} \frac{w_i + w_K}{w_i w_K}, & i = j \\ \frac{1}{w_K}, & i \neq j \end{cases}$$

and Jacobian matrix

$$\boldsymbol{J_w} = \frac{1}{w_K} \begin{pmatrix} 1 + w_K/w_1 & 1 & \ldots & 1 \\ 1 & 1 + w_K/w_2 & \ldots & 1 \\ \vdots & & & \vdots \\ 1 & \ldots & 1 & 1 + w_K/w_{K-1} \end{pmatrix}$$

with determinant

$$|\boldsymbol{J_w}| = \prod_{j=1}^{K} w_j^{-1}.$$

For τ, finally, we have the restriction $\tau_{min} < \tau < \tau_{max}$ and we can transform τ to the real line by

$$\widetilde{\tau} = g_\tau(\tau) = \ln\left(\frac{\tau - \tau_{min}}{\tau_{max} - \tau}\right)$$

with Jacobian

$$\frac{\partial \widetilde{\tau}}{\partial \tau} = \frac{\tau_{max} - \tau_{min}}{(\tau - \tau_{min})(\tau_{max} - \tau)}.$$

Transform the MCMC output from $\boldsymbol{\theta}^{(i)} = \left(\boldsymbol{\Gamma}^{(i)}, \boldsymbol{\psi}^{(i)}, \boldsymbol{w}^{(i)}, \tau^{(i)}\right)$ to $\widetilde{\boldsymbol{\theta}}^{(i)} = \left(\boldsymbol{\Gamma}^{(i)}, \widetilde{\boldsymbol{\psi}}^{(i)}, \widetilde{\boldsymbol{w}}^{(i)}, \widetilde{\tau}^{(i)}\right)$ and let $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{S}$ be the sample mean and variance-covariance matrix of $\widetilde{\boldsymbol{\theta}}$. Next, let

$$f\left(\widetilde{\boldsymbol{\theta}}\right) \propto \exp\left[-\frac{1}{2}\left(\widetilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\right)' \boldsymbol{S}^{-1}\left(\widetilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\right)\right] I_{\chi^2_\alpha}$$

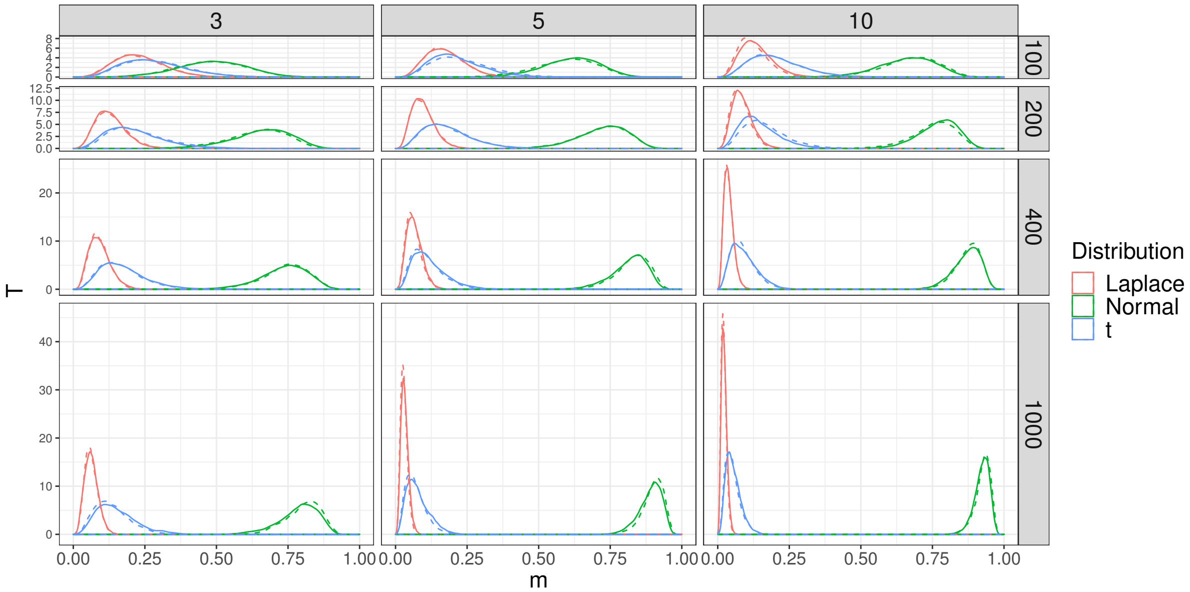
where $I_{\chi^2_\alpha}$ is the indicator function for the set $\left\{\widetilde{\boldsymbol{\theta}} : \left(\widetilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\right)' \boldsymbol{S}^{-1}\left(\widetilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\right) \leq \chi^2_\alpha(n)\right\}$ and $\chi^2_\alpha(n)$ is the α quantile of the χ^2 distribution with n degrees of freedom (the dimension of $\boldsymbol{\theta}$). Our modified harmonic mean estimator of the marginal likelihood is then

$$\widehat{m}\left(\boldsymbol{Y}\right) = \left[\frac{1}{R}\sum_{i=1}^{R} \frac{f\left(\widetilde{\boldsymbol{\theta}}^{(i)}\right)}{m\left(\boldsymbol{Y}|\boldsymbol{\theta}^{(i)}\right)\pi\left(\boldsymbol{\theta}^{(i)}\right)|\boldsymbol{J}_{\widetilde{\boldsymbol{\theta}}}|}\right]^{-1}$$

with $|\boldsymbol{J}_{\widetilde{\boldsymbol{\theta}}}| = |\boldsymbol{J}_{\widetilde{\boldsymbol{\psi}}}||\boldsymbol{J_w}|^{-1}|\boldsymbol{J}_\tau|^{-1}$.

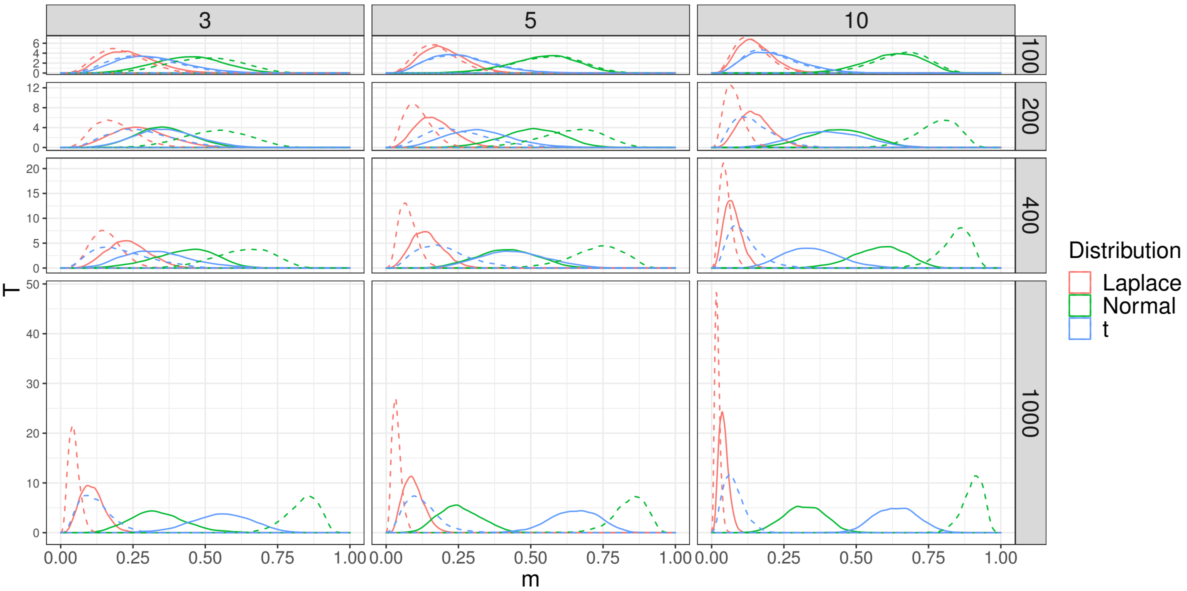
B Additional figures

Figure 16 Posterior distributions of the weights, normally distributed data



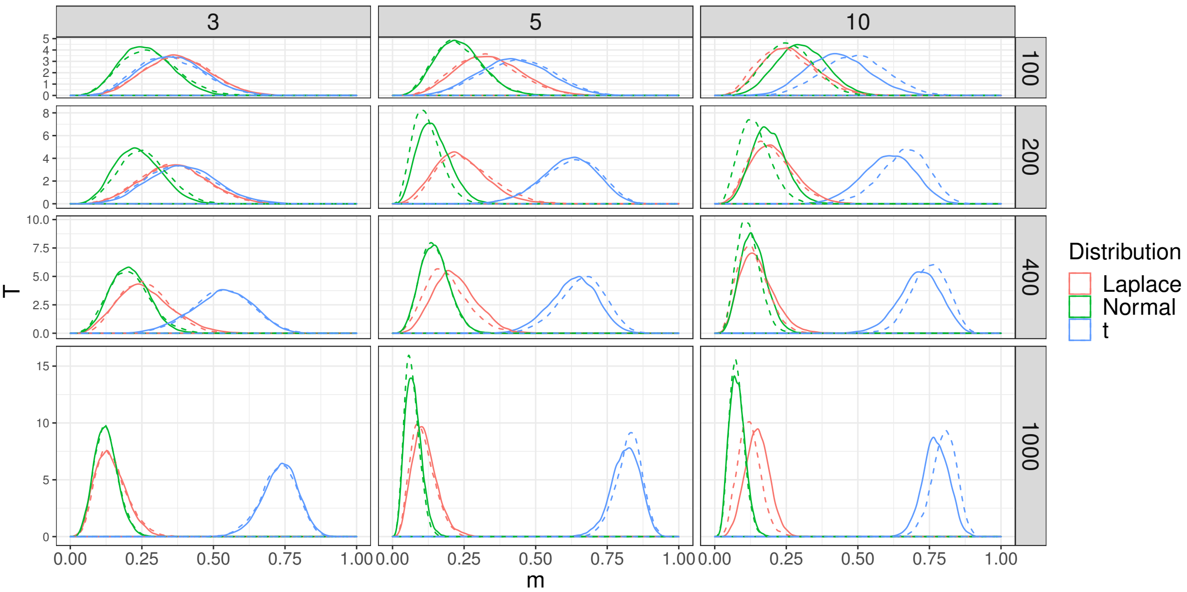
The true distribution of the error terms is generated from a multivariate standard normal distribution. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 17 Posterior distributions of the weights, normally distributed data with stochastic volatility



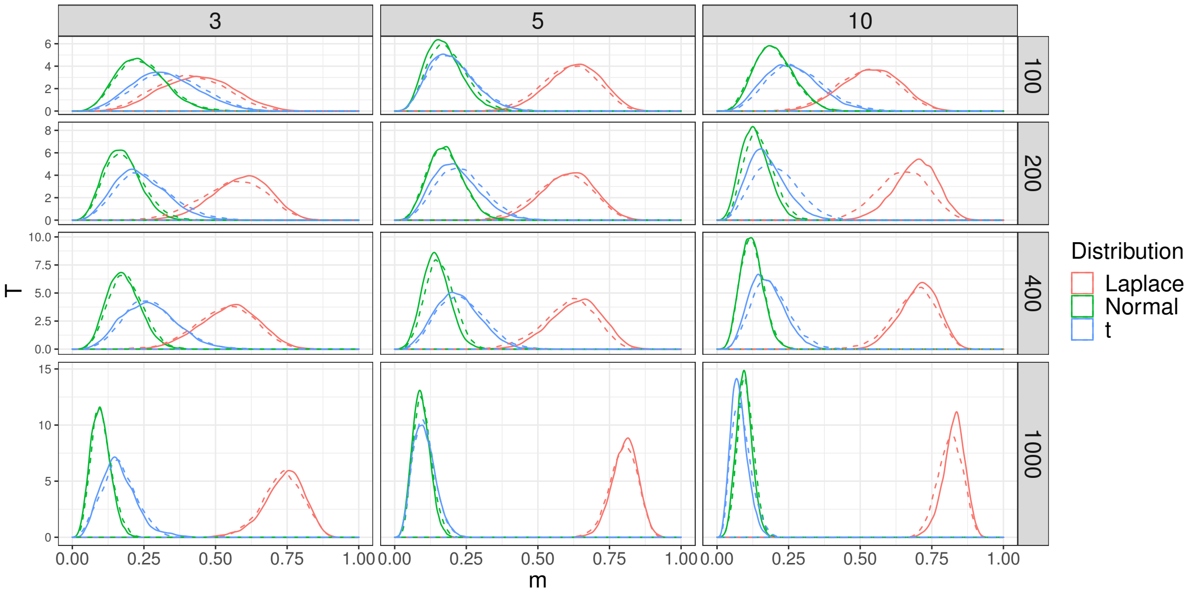
The true distribution of the error terms is generated from a multivariate normal distribution with stochastic volatility. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 18 Posterior distributions of the weights, t-distributed data with 3 degrees of freedom



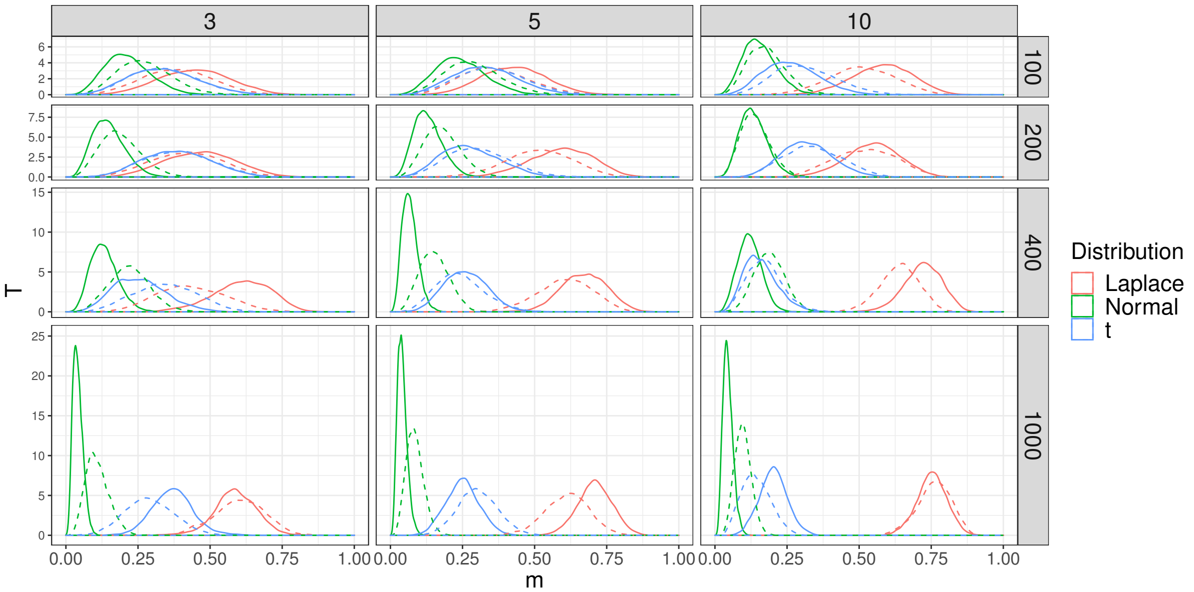
The true distribution of the error terms is generated from a multivariate standard t distribution with 3 degrees of freedom. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 19 Posterior distributions of the weights, Laplace distributed data



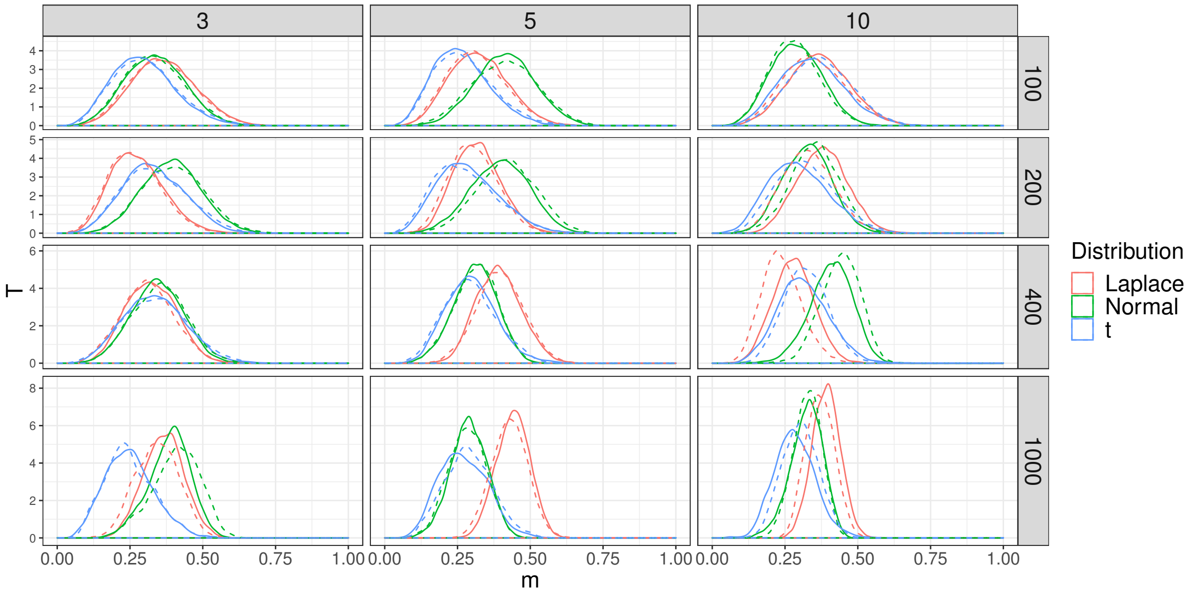
The true distribution of the error terms is generated from a multivariate standard Laplace distribution. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 20 Posterior distributions of the weights, Laplace distributed data with stochastic volatility



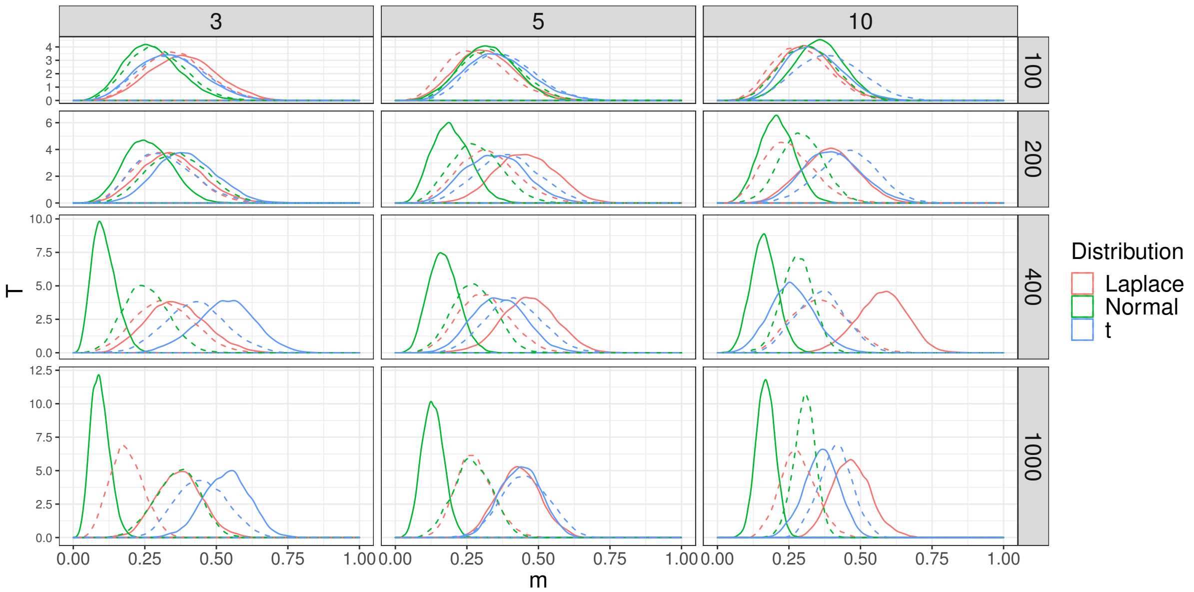
The true distribution of the error terms is generated from a multivariate Laplace distribution with stochastic volatility. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 21 Posterior distributions of the weights, mixture distributed data



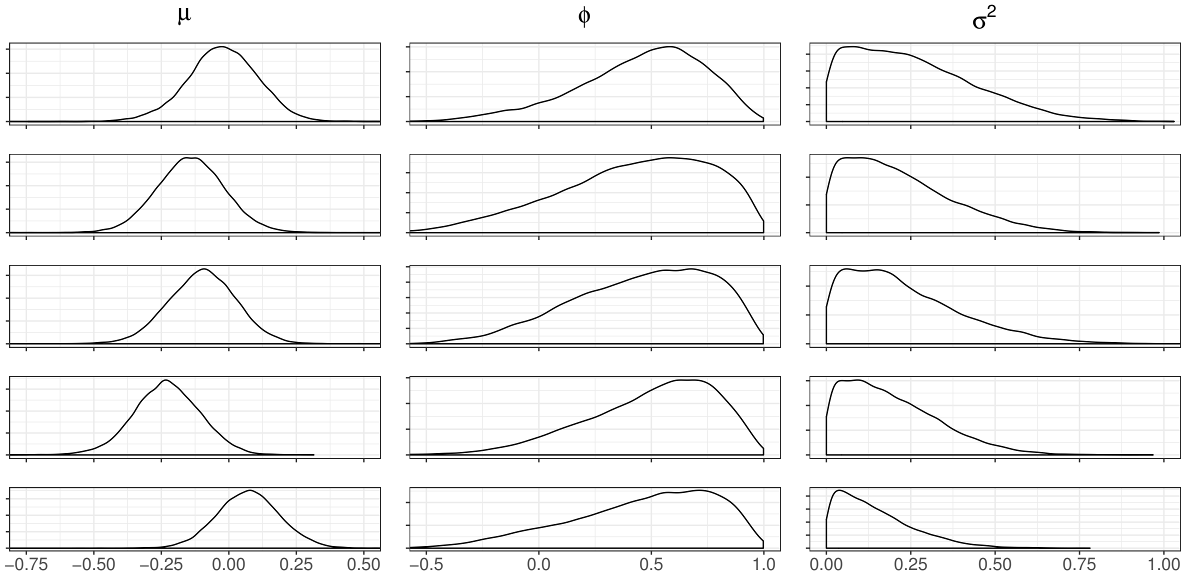
The true distribution of the error terms is generated from the mixture of a multivariate standard normal, t with 5 degrees of freedom and Laplace distributions such that $\boldsymbol{w} = (1/3, 1/3, 1/3)$. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 22 Posterior distributions of the weights, mixture distributed data with stochastic volatility



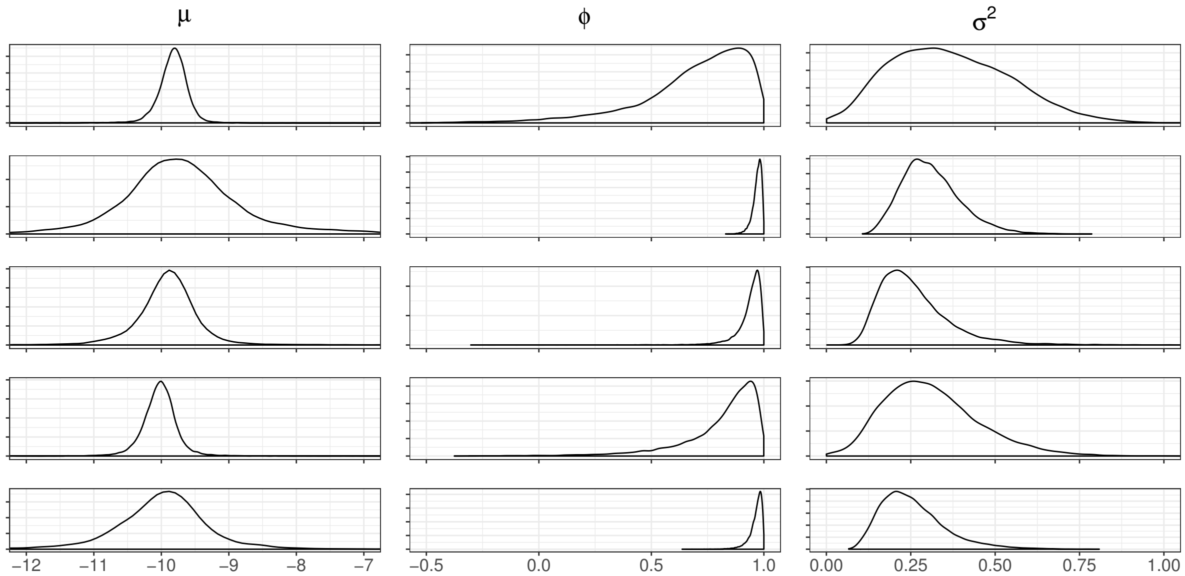
The true distribution of the error terms is generated from a mixture of multivariate normal, t with 5 degrees of freedom and Laplace distributions with stochastic volatility such that $\boldsymbol{w} = (1/3, 1/3, 1/3)$. The posterior distributions of the weights that are obtained via MCMC-sampler without stochastic volatility are plotted as solid lines, while the posterior distributions of the weights that are obtained via MCMC-sampler with stochastic volatility are plotted as dashed lines. Several dimensions $m \in \{3, 5, 10\}$ and observations $T \in \{100, 200, 400, 1000\}$ are considered.

Figure 23 Posterior distributions of volatility parameters, normally distributed data



The true distribution of the error terms is generated from a multivariate standard normal distribution. The dimension m is taken to be 5, while the number of observations T is equal to 200.

Figure 24 Posterior distributions of volatility parameters, normally distributed data with stochastic volatility



The true distribution of the error terms is generated from a multivariate standard normal distribution with stochastic volatility. The dimension m is taken to be 5, while the number of observations T is equal to 200.

C Data

C.1 A small VAR

The data is obtained from the OECD Main Economic Indicators (`https://stats.oecd.org/`).

- Inflation is based on the personal consumption price index (CPI) and calculated as

$$100 \times (CPI_t - CPI_{t-12})/CPI_{t-12}.$$

  Where CPI is the series CPI: All items (COICOP 1-12).

- Unemployment is the seasonally adjusted unemployment rate (Harmonised unemployment rate (monthly), Total, All persons).

- The interest rate is the 3-month or 90 day interbank rate.

Table 4 Sample period for countries

Country	Start	End
Austria	1993M1	2020M1
Canada	1956M1	2020M1
Denmark	1987M1	2020M1
Finland	1988M1	2020M1
France	1983M1	2020M1
Germany	1991M1	2020M1
Italy	1983M1	2020M1
Japan	2002M4	2020M1
Korea	1991M1	2020M1
Netherlands	1983M1	2020M1
Norway	1989M1	2019M12
Spain	1986M4	2020M1
Sweden	1983M1	2020M1
United Kingdom	1986M1	2019M11
United States	1964M6	2020M2

C.2 A medium sized VAR

The data is downloaded from FRED (`https://fred.stlouisfed.org`).

- The real GDP is evaluated via the level of real gross domestic product (GDPC1) and the quarterly average of the Civilian Non-institutional Population (CNP16OV) which is normalized so that its 1992Q3 value is one, and is given by

$$100 \times \log\left(GDP_t/POP_t\right).$$

- The real consumption is based on the the level of personal consumption (PCEC), the level of the GDP price deflator (GDPDEF), and the normalized quarterly average of the Civilian Non-institutional Population (CNP16OV), and is evaluated as

$$100 \times \log\left(CONS_t/(GDPP_t \times POP_t)\right).$$

- The real investment is calculated using the level of fixed private investment (FPI), the level of the GDP price deflator (GDPDEF), and the normalized quarterly average of the Civilian Non-institutional Population (CNP16OV), and is evaluated as
$$100 \times \log\left(INV_t/(GDPP_t \times POP_t)\right).$$
- The real wage is evaluated via the compensation per hour for the nonfarm business sector (COMPNFB) and the level of the GDP price deflator (GDPDEF), and is given by
$$100 \times \log(W_t/GDPP_t).$$
- Hours worked are computed using the index of average weekly nonfarm business hours (PRS85006023), the normalized quarterly average of the Civilian Non-institutional Population (CNP16OV), the number of employed civilians (CE16OV) which is normalized so that its 1992Q3 value is 1, and are given by
$$100 \times \log\left(HOURS_t \times EMP_t/POP_t\right).$$
- The inflation rate is based on the GDP price deflator (GDPDEF) and calculated as
$$100 \times \log\left(GDPP_t/GDPP_{t-1}\right).$$
- The federal funds rate is based on the effective federal funds rate (FEDFUNDS) and evaluated as $FFR_t/4$.

References

Chan, J. C. C. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter vars with stochastic volatility. Journal of Applied Econometrics, 33:509–532.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association, 96:270–281.

Chiu, C.-W. J., Mumtaz, H., and Pinter, G. (2017). Forecasting with var models: Fat tails and stochastic volatility. International Journal of Forecasting, 33(4):1124–1143.

Cogley, T. and Sargent, T. J. (2002). Evolving post-world war II U.S. inflation dynamics. In Bernanke, B. S. and Rogoff, K. S., editors, NBER Macroeconomics Annual 2001, Volume 16, pages 331–388. National Bureau of Economic Research, Inc.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post wwii us. Review of Economic Dynamics, 8:262–302.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypothesis on normal location parameters. The Annals of Mathematical Statistics, 42:204–223.

Doan, T., Litterman, R. B., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. Econometric Reviews, 3:1–144.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. Journal of Econometrics, 154(1):85–100.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. Journal of the Royal Statistical Society, Ser. B, 56:501–514.

Geweke, J. (1999). Using simulation methods for bayesian econometric models: Inference, development and communication. Econometric Reviews, 18:1–126. with discussion.

Gupta, A., Nagar, D., and Bodnar, T. (2013). Elliptically Contoured Models in Statistics and Portfolio Theory. Springer.

Highfield, R. A. (1987). Forecasting with Bayesian State Space Models. PhD thesis, Graduate School of Business, University of Chicago.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in bayesian VAR-models. Journal of Applied Econometrics, 12:99–132.

Karlsson, S. (2013). Handbook of Economic Forecasting, volume 2, chapter Forecasting with Bayesian Vector Autoregression, pages 791–897. Elsevier.

Kastner, G. (2016). Dealing with stochastic volatility in time series using the r package stochvol. Journal of Statistical Software, 69(5):1–30.

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

Koop, G. and Potter, S. M. (1999). Bayes factors and nonlinearity: Evidence from economic time seriess. Journal of Econometrics, 88:251–281.

Lee, N., Choi, H., and Kim, S.-H. (2016). Bayes shrinkage estimation for high-dimensional var models with scale mixture of normal distributions for noise. Computational Statistics & Data Analysis, 101(C):250–276.

Linton, O. and McCrorie, J. R. (1995). Differentiation of an exponential matrix function. Econometric Theory, 11(5):1182–1185.

Ni, S. and Sun, D. (2005). Bayesian estimates for vector autoregressive models. Journal of Business & Economic Statistics, 23:105–117.

Panagiotelis, A. and Smith, M. (2008). Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. International Journal of Forecasting, 24(4):710 – 727.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. The Review of Economic Studies, 72(3):821–852.

Sims, C. A. (1993). A nine-variable probabalistic macroeconomic forecasting model. In Stock, J. H. and Watson, M. W., editors, Business Cycles, Indicators and Forecasting, pages 179–204. University of Chicago Press.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. International Econom Review, 39:949–968.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. The American Economic Review, 97(3):586–606.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. Econometrica, 65:59–73.