Article

Triple the Gamma—A Unifying Shrinkage Prior for Variance and Variable Selection in Sparse State Space and TVP Models

Annalisa Cadonna, Sylvia Frühwirth-Schnatter * and Peter Knaus**

Department of Finance, Accounting and Statistics, WU Vienna University of Economics and Business, 1020 Vienna, Austria; annalisa.cadonna@wu.ac.at (A.C.); peter.knaus@wu.ac.at (P.K.)

* Correspondence: sfruehwi@wu.ac.at

check for updates

Abstract: Time-varying parameter (TVP) models are very flexible in capturing gradual changes in the effect of explanatory variables on the outcome variable. However, in particular when the number of explanatory variables is large, there is a known risk of overfitting and poor predictive performance, since the effect of some explanatory variables is constant over time. We propose a new prior for variance shrinkage in TVP models, called triple gamma. The triple gamma prior encompasses a number of priors that have been suggested previously, such as the Bayesian Lasso, the double gamma prior and the Horseshoe prior. We present the desirable properties of such a prior and its relationship to Bayesian Model Averaging for variance selection. The features of the triple gamma prior are then illustrated in the context of time varying parameter vector autoregressive models, both for simulated dataset and for a series of macroeconomics variables in the Euro Area.

1. Introduction

Model selection in a high-dimensional setting is a common challenge in statistical and econometric inference. The introduction of Bayesian model averaging (BMA) techniques in the statistical literature (Brown et al. 2002; Cottet et al. 2008; Raftery et al. 1997) has led to many interesting applications, see, among others, (Frühwirth-Schnatter and Tüchler 2008; Kleijn and van Dijk 2006; Koop and Potter 2004; Sala-i-Martin et al. 2004) for early references in econometrics.

Selecting explanatory variables for possibly very high-dimensional regression problems though shrinkage priors is an attractive alternative to BMA which relies on discrete mixture priors, see Bhadra et al. (2019) for an excellent review. There is a vast and growing literature on shrinkage priors for regression problems that focuses on the following aspects. First, how to choose sensible priors for high-dimensional model selection problems in a Bayesian framework, second, how to design efficient algorithms to cope with the associated computational challenges and third, to investigate, both from a theoretical and a practical viewpoint, how such priors perform in high-dimensional problems.

A striking duality exists in this very active area between Bayesian and traditional approaches. For many shrinkage priors, the mode of the posterior distribution obtained in a Bayesian analysis can be regarded as a point estimate from a regularization approach, see Fahrmeir et al. (2010) and Polson and Scott (2012a). One such example is the popular Lasso (Tibshirani 1996) which is equivalent to a double-exponential shrinkage prior in a Bayesian context (Park and Casella 2008). However, the two approaches differ when it comes to selecting penalty parameters that impact the sparsity of the

solution. One advantage of the Bayesian framework in this context is that the penalty parameters are considered to be unknown hyperparameters which can be learned from the data. Such "global-local" shrinkage priors (Polson and Scott 2011) adjust to the overall degree of sparsity that is required in a specific application through a global shrinkage parameter and separate signal from noise through local, individual shrinkage parameters.

While the inclusion of potentially many explanatory variables though shrinkage priors in regression models is addressed in a vast literature, the use of shrinkage priors for more general econometric models in time series analysis, such as state space models and time-varying parameter (TVP) models is, in comparison, less well-studied. Sparsity in the context of such models refers to the presence of a few large variances among many (nearly) zero variances in the latent state processes that drive the observed time series data. A common goal in this setting is to recover a few dynamic states, driven by such a state space model, among many (nearly) constant coefficients. As shown by Frühwirth-Schnatter and Wagner (2010), this variance selection problem can be cast into a variable selection problem in the non-centered parametrization of a state space model. Once this link has been established, shrinkage priors that are known to perform well in high-dimensional regression problems can be applied to variance selection in state space models, as demonstrated for the Lasso (Belmonte et al. 2014) and the normal-gamma (Bitto and Frühwirth-Schnatter 2019; Griffin and Brown 2017).

Despite this already existing variety, we introduce a new shrinkage prior for variance selection in sparse state space and TVP models in the present paper called triple gamma prior, as it has a representation involving three gamma distributions. This prior can be related to various shrinkage priors that were found to be useful for high-dimensional regression problems, such as the generalized beta mixture prior (Armagan et al. 2011), and contains the popular Horseshoe prior (Carvalho et al. 2009 2010) as a special case. Furthermore, the half-*t* and the half Cauchy (Gelman 2006; Polson and Scott 2012b), suggested as robust alternatives to the inverse gamma distribution for variance parameters in hierarchical models, as well as the Lasso and the double gamma, are special cases of the triple gamma. In this context, the triple gamma can also be regarded as an extension of the scaled beta2 distribution (Pérez et al. 2017).

Among Bayesian shrinkage priors, usually a clear distinction is made between two-group mixture or spike-and-slab priors and continuous shrinkage priors, of which the triple gamma is a special case. An important contribution of the present paper is to show that the triple gamma provides a bridge between these two approaches and has the following property which is favourable both in sparse and dense situations. One of the hyperparameters allows high concentration over the region in the shrinkage profile that is relevant for shrinking noise, while the other hyperparameter allows high concentration over the region that prevents overshrinking of signals. This allows the triple gamma prior to exhibit behavior that very much resembles Bayesian model averaging based on discrete spike-and-slab priors, with a strong prior concentration at the corner solutions where some of the variances are nearly zero. While this is reminiscent of the Horseshoe prior, the shrinkage profile induced by the triple gamma is more flexible than that of a Horseshoe. Thanks to the estimation of the hyperparemters, it is not constrained to be symmetric around one half, enabling adaption to varying degrees of sparsity in the data.

The triple gamma prior also scores well from a computational perspective. While exploring the full posterior distribution for spike-and-slab priors leads to computational challenges due to the combinatorial complexity of the model space, Bayesian inference based on Markov chain Monte Carlo (MCMC) methods is straightforward for continuous shrinkage priors, exploiting their Gaussian-scale mixture representation (Bitto and Frühwirth-Schnatter 2019; Makalic and Schmidt 2016). An extension of these schemes to the triple gamma prior is fairly straightforward.

We will study the empirical performance of the triple gamma for a challenging setting in econometric time series analysis, namely for time-varying parameter vector autoregressive models with stochastic volatility (TVP-VAR-SV models). Since the influential paper of Primiceri (2005)

(see Del Negro and Primiceri (2015) for a corrigendum), this model has become a benchmark for analyzing relationships between macroeconomic variables that evolve over time, see Nakajima (2011), Koop and Korobilis (2013), Eisenstat et al. (2014), Chan and Eisenstat (2016), Feldkircher et al. (2017) and Carriero et al. (2019), among many others. Due to the high dimensionality of the time-varying parameters, even for moderately sized systems, shrinkage priors such as the triple gamma prior are instrumental for efficient inference.

The rest of the paper is organized as follows—in Section 2, we define the triple gamma prior and discuss some of its properties. The close relationship between the triple gamma and spike-and-slab priors applied in a BMA context is investigated in Section 3.2. Section 4 introduces an efficient MCMC scheme and Section 5 provides applications to TVP-VAR-SV models. Section 6 concludes the paper.

2. The Triple Gamma as a Prior for Variance Parameters

2.1. Motivation and Definition

To motivate the triple gamma prior, consider the state space form of a TVP model for a univariate time series y_t. For $t = 1, \ldots, T$, we have that

$$
\begin{aligned}
\boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \mathbf{w}_t, & \mathbf{w}_t &\sim \mathcal{N}_d\left(\mathbf{0}, \mathbf{Q}\right), \\
y_t &= \mathbf{x}_t \boldsymbol{\beta}_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}\left(0, \sigma_t^2\right),
\end{aligned}
\tag{1}
$$

where $\mathbf{Q} = \text{Diag}\left(\theta_1, \ldots, \theta_d\right)$ and the initial value of the state process follows a normal distribution, $\boldsymbol{\beta}_0 \sim \mathcal{N}_d\left(\boldsymbol{\beta}, \mathbf{Q}\right)$, with initial mean $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_d\right)^\top$. $\mathbf{x}_t = \left(x_{t1}, \ldots, x_{td}\right)$ is a d-dimensional row vector containing the explanatory variables at time t. The variables x_{tj} can be exogenous control variables and/or be equal to lagged values of y_t. Usually, one of the variables, say x_{t1}, corresponds to the intercept, but an intercept need not be present. This approach can be straightforwardly adapted to the multivariate case as for the TVP-VAR-SV model that will be considered in Section 5.

The error variance σ_t^2 in the observation equation is either homoscedastic ($\sigma_t^2 \equiv \sigma^2$ for all $t = 1, \ldots, T$) or follows a stochastic volatility (SV) specification (Jacquier et al. 1994), where the log volatility $h_t = \log \sigma_t^2$ follows an AR(1) process. Specifically,

$$
h_t | h_{t-1}, \mu, \phi, \sigma_\eta^2 \sim \mathcal{N}\left(\mu + \phi(h_{t-1} - \mu), \sigma_\eta^2\right).
\tag{2}
$$

For Bayesian inference, priors have to be chosen for the unknown variances $\theta_1, \ldots, \theta_d$ and the unknown initial means $\beta_1, \ldots, \beta_d$. In order to shrink dynamic coefficients to static ones and, in this way, avoid overfitting, a shrinkage prior is placed on θ_j that puts a lot of prior mass close to zero. One such prior is the double gamma prior, employed recently by Bitto and Frühwirth-Schnatter (2019). The double gamma prior can be expressed as a scale-mixture of gamma distributions, with the following hierarchical representation:

$$
\theta_j | \xi_j^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2\xi_j^2}\right), \quad \xi_j^2 | a^\xi, \kappa_B^2 \sim \mathcal{G}\left(a^\xi, \frac{a^\xi \kappa_B^2}{2}\right).
\tag{3}
$$

In the double gamma prior, each innovation variance θ_j is mixed over its own scale parameter ξ_j^2, each of which has an independent gamma distribution, with a common hyperparameter κ_B^2. Moreover, the ξ_j^2's play the role of local (component specific) shrinkage parameters, while the parameter κ_B^2 is a (common) global shrinkage parameter.

We propose an extension of the double gamma prior to a triple gamma prior, where another layer is added to the hierarchy:

$$\theta_j|\xi_j^2 \sim \mathcal{G}\left(\frac{1}{2},\frac{1}{2\xi_j^2}\right), \quad \xi_j^2|a^\xi,\kappa_j^2 \sim \mathcal{G}\left(a^\xi,\frac{a^\xi\kappa_j^2}{2}\right), \quad \kappa_j^2|c^\xi,\kappa_B^2 \sim \mathcal{G}\left(c^\xi,\frac{c^\xi}{\kappa_B^2}\right). \tag{4}$$

The main difference with the double gamma prior is that the prior scale of the ξ_j^2's is not identical, but each ξ_j^2 depends on its component specific scale κ_j^2. We will show in Section 2.2 that the triple gamma prior can be represented as a global-local shrinkage prior in the sense of Polson and Scott (2012a) where the local shrinkage parameters ξ_j^2 arise from an F $\left(2a^\xi, 2c^\xi\right)$ distribution. Hence, the triple gamma prior contains the Horseshoe prior and many other well-known shrinkage priors as special cases, as will be discussed in Section 2.3.

The shrinkage behaviour of the triple gamma prior becomes even more apparent when we rewrite model (1) in the non-centered parametrization introduced in Frühwirth-Schnatter and Wagner (2010):

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_t &= \tilde{\boldsymbol{\beta}}_{t-1} + \tilde{\mathbf{w}}_t, \qquad \tilde{\mathbf{w}}_t \sim \mathcal{N}_d\left(\mathbf{0},\mathbf{I}_d\right), \\ y_t &= \mathbf{x}_t\boldsymbol{\beta} + \mathbf{x}_t\text{Diag}\left(\sqrt{\theta_1},\ldots,\sqrt{\theta_d}\right)\tilde{\boldsymbol{\beta}}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\sigma_t^2\right),\end{aligned} \tag{5}$$

with $\tilde{\boldsymbol{\beta}}_0 \sim \mathcal{N}_d\left(\mathbf{0},\mathbf{I}_d\right)$, where $\mathbf{I}_d$ is the d-dimensional identity matrix. Both representations are equivalent and we can specify a prior either on the variances θ_j in (1) or the scale parameters $\sqrt{\theta_j}$ in (5). Using the fact that $\theta_j/\xi_j^2 \sim \chi_1^2$ and the χ_1^2-distribution can be represented as $\chi_1^2 = Z_j^2$, where $Z_j \sim \mathcal{N}\left(0,1\right)$ follows a standard normal distribution, we can match prior (4) to the non-centered parametrization (5). This yields

$$\sqrt{\theta_j}|\xi_j^2 \sim \mathcal{N}\left(0,\xi_j^2\right), \quad \xi_j^2|a^\xi,\kappa_j^2 \sim \mathcal{G}\left(a^\xi,\frac{a^\xi\kappa_j^2}{2}\right), \quad \kappa_j^2|c^\xi,\kappa_B^2 \sim \mathcal{G}\left(c^\xi,\frac{c^\xi}{\kappa_B^2}\right). \tag{6}$$

In (6), we could force $\sqrt{\theta_j}$ to take on only positive values, however, we do not impose such a constraint and allow $\sqrt{\theta_j}$ to take on negative values. Since the half-normal $\sqrt{\theta_j} \sim \mathcal{N}\left(0,\xi_j^2\right) I\{\sqrt{\theta_j} > 0\}$ also implies that $\theta_j \sim \xi_j^2\chi_1^2$, the question arises whether the negative half is of importance. Whenever inference is performed under the non-centered parametrization (5), as is done in Section 4, restricting the prior to the positive half will lead to automatic truncation of the full conditional posterior $p(\sqrt{\theta_j}|\tilde{\boldsymbol{\beta}}_0,\ldots,\tilde{\boldsymbol{\beta}}_T,\mathbf{y},\cdot)$ to the positive part during MCMC sampling. If the positive and the negative mode of the marginal posterior $p(\sqrt{\theta_j}|\mathbf{y})$ are well-separated, then this will not matter. However, if the true value of θ_j is close or equal to zero and $p(\sqrt{\theta_j}|\mathbf{y})$ is concentrated at zero, this truncation will introduce a bias, because the negative half is not accounted for.

Interestingly, prior (6) is related to the so-called normal-gamma-gamma prior consider by Griffin and Brown (2017) in the context of defining hierarchical shrinkage priors for regression models. This relation is helpful in choosing a prior on the fixed coefficients $\beta_1,\ldots,\beta_d$. To allow shrinkage of these coefficients toward insignificant ones in a TVP model, we extend Bitto and Frühwirth-Schnatter (2019) further by assuming such a normal-gamma-gamma prior on $\beta_1,\ldots,\beta_d$:

$$\beta_j|\tau_j^2 \sim \mathcal{N}\left(0,\tau_j^2\right), \quad \tau_j^2|a^\tau,\lambda_j^2 \sim \mathcal{G}\left(a^\tau,\frac{a^\tau\lambda_j^2}{2}\right), \quad \lambda_j^2|c^\tau,\lambda_B^2 \sim \mathcal{G}\left(c^\tau,\frac{c^\tau}{\lambda_B^2}\right). \tag{7}$$

In Section 2.4, we will discuss hierarchical versions of both priors, by putting a hyperprior on the parameters κ_B^2, λ_B^2, a^ξ, a^τ, c^ξ, and c^τ.

2.2. Properties of the Triple Gamma Prior

In this section, we study the mathematical properties of the triple gamma prior. It is shown in Theorem 1 that the triple gamma prior is a global-local shrinkage prior where the local shrinkage parameters arise from the $F\left(2a^{\tilde{\xi}}, 2c^{\tilde{\xi}}\right)$ distribution. Furthermore, a closed form of the marginal shrinkage prior $p(\sqrt{\theta_j}|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}})$ is given in Theorem 1, which is proven in Appendix A.

Theorem 1. *For the triple gamma prior defined in (4), with $a^{\tilde{\xi}} > 0$ and $c^{\tilde{\xi}} > 0$, the following holds:*

(a) *It has following representation as a local-global shrinkage prior:*

$$\sqrt{\theta_j}|\psi_j^2, \kappa_B^2 \sim \mathcal{N}\left(0, \frac{2}{\kappa_B^2}\psi_j^2\right), \quad \psi_j^2|a^{\tilde{\xi}}, c^{\tilde{\xi}} \sim F\left(2a^{\tilde{\xi}}, 2c^{\tilde{\xi}}\right). \tag{8}$$

(b) *The marginal prior $p(\sqrt{\theta_j}|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}})$ takes the following form with $\phi^{\tilde{\xi}} = \frac{2c^{\tilde{\xi}}}{\kappa_B^2 a^{\tilde{\xi}}}$,*

$$p(\sqrt{\theta_j}|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}}) = \frac{\Gamma(c^{\tilde{\xi}} + \frac{1}{2})}{\sqrt{2\pi\phi^{\tilde{\xi}}}B(a^{\tilde{\xi}}, c^{\tilde{\xi}})}U\left(c^{\tilde{\xi}} + \frac{1}{2}, \frac{3}{2} - a^{\tilde{\xi}}, \frac{\theta_j}{2\phi^{\tilde{\xi}}}\right), \tag{9}$$

where $U(a, b, z)$ is the confluent hyper-geometric function of the second kind:

$$U(a, b, z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt}t^{a-1}(1+t)^{b-a-1}dt.$$

In Figure 1, we can see the marginal prior distribution of $\sqrt{\theta_j}$ under the triple gamma prior $a^{\tilde{\xi}} = c^{\tilde{\xi}} = 0.1$ and under other well-known shrinkage priors which are special cases of the triple gamma, see Table 1. Theorem 1 also allows us to give a closed form for the prior $p(\theta_j|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}}) = p(\sqrt{\theta_j}|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}})/\sqrt{\theta_j}$.[1]

Global-local shrinkage priors are typically compared in terms of the concentration around the origin and the tail behaviour. For the triple gamma prior $p(\sqrt{\theta_j}|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}})$, the two shape parameters $a^{\tilde{\xi}}$ and $c^{\tilde{\xi}}$ play a crucial role in this respect, see Theorem 2 which is proven in Appendix A.

Theorem 2. *The triple gamma prior (9) satisfies the following:*

(a) *For $0 < a^{\tilde{\xi}} < 0.5$ and small values of $\sqrt{\theta_j}$,*

$$p(\sqrt{\theta_j}|\phi^{\tilde{\xi}}, a^{\tilde{\xi}}, c^{\tilde{\xi}}) = \frac{\Gamma(\frac{1}{2} - a^{\tilde{\xi}})}{\sqrt{\pi}(2\phi^{\tilde{\xi}})^{a^{\tilde{\xi}}}B(a^{\tilde{\xi}}, c^{\tilde{\xi}})}\left(\frac{1}{\sqrt{\theta_j}}\right)^{1-2a^{\tilde{\xi}}} + O(1).$$

---

[1] Let $f_{\sqrt{\theta_j}}(x)$ and $F_{\sqrt{\theta_j}}(x)$ be, respectively, the pdf and cdf of the random variable $\sqrt{\theta_j}$. The cdf $F_{\theta_j}(x)$ of the random variable θ_j is given by

$$F_{\theta_j}(x) = \Pr(\theta_j \le x) = \Pr(-\sqrt{x} \le \sqrt{\theta_j} \le \sqrt{x}) = F_{\sqrt{\theta_j}}(\sqrt{x}) - F_{\sqrt{\theta_j}}(-\sqrt{x}) = 2F_{\sqrt{\theta_j}}(\sqrt{x}),$$

since $f_{\sqrt{\theta_j}}(x)$ is symmetric around 0. The pdf $f_{\theta_j}(x)$ is obtained by taking the first derivative of $F_{\theta_j}(x)$ with respect to x:

$$f_{\theta_j}(x) = \frac{d F_{\theta_j}(x)}{d x} = f_{\sqrt{\theta_j}}(\sqrt{x})/\sqrt{x}.$$

(b)   For $a^{\xi} = 0.5$ and small values of $\sqrt{\theta_j}$,

$$p(\sqrt{\theta_j}|\phi^{\xi}, a^{\xi}, c^{\xi}) = \frac{1}{\sqrt{2\pi\phi^{\xi}}B(0.5, c^{\xi})}\left(-\log\theta_j + \log(2\phi^{\xi}) - \psi(c^{\xi} + 0.5)\right) + O(|\theta_j\log\theta_j|),$$

where $\psi(\cdot)$ is the digamma function.

(c)   For $a^{\xi} > 0.5$,

$$\lim_{\sqrt{\theta_j} \to 0} p(\sqrt{\theta_j}|\phi^{\xi}, a^{\xi}, c^{\xi}) = \frac{\Gamma(c^{\xi} + \frac{1}{2})\Gamma(a^{\xi} - \frac{1}{2})}{\sqrt{2\pi\phi^{\xi}}\Gamma(c^{\xi})\Gamma(a^{\xi})}.$$

(d)   As $\sqrt{\theta_j} \to \infty$,

$$p(\sqrt{\theta_j}|\phi^{\xi}, a^{\xi}, c^{\xi}) = \frac{\Gamma(c^{\xi} + \frac{1}{2})(2\phi^{\xi})^{c^{\xi}}}{\sqrt{\pi}B(a^{\xi}, c^{\xi})}\left(\frac{1}{\sqrt{\theta_j}}\right)^{2c^{\xi}+1}\left[1 + O\left(\frac{1}{\theta_j}\right)\right].$$

From Theorem 2, Part (a) and (b), we find that the triple gamma prior $p(\sqrt{\theta_j}|\phi^{\xi}, a^{\xi}, c^{\xi})$ has a pole at the origin, if $a^{\xi} \leq 0.5$. According to Part (a), the pole is more pronounced, the closer a^{ξ} gets to 0. For $a^{\xi} > 0.5$, we find from Part (c) that $p(\sqrt{\theta_j}|\phi^{\xi}, a^{\xi}, c^{\xi})$ is bounded at zero by a positive upper bound which is finite, as long as $0 < c^{\xi} < \infty$. Part (d) shows that the triple gamma prior $p(\sqrt{\theta_j}|\phi^{\xi}, a^{\xi}, c^{\xi})$ has polynomial tails, with the shape parameter c^{ξ} controlling the tail index. Prior moments $E((\sqrt{\theta_j})^k|\phi^{\xi}, a^{\xi}, c^{\xi})$ exist up to $k < 2c^{\xi}$. Hence, the triple gamma prior has no finite moments for $c^{\xi} < 1/2$.

Finally, additional useful representations of the triple gamma prior as a global-local shrinkage prior are summarized in Lemma 1 which is proven in Appendix A. Representation (a) shows that the triple gamma is an extension of the double gamma prior where the Gaussian prior $\sqrt{\theta_j}|\xi_j^2 \sim \mathcal{N}(0, \xi_j^2)$ is substituted by a heavier-tailed Student-t prior, making the prior more robust to large values of $\sqrt{\theta_j}$. Representation (b) and (c) will be useful for MCMC inference in Section 4. Representations (c) and (d) show that for a triple gamma prior with finite a^{ξ} and c^{ξ}, ϕ^{ξ} acts as a global shrinkage parameter, in addition to $2/\kappa_B^2$.
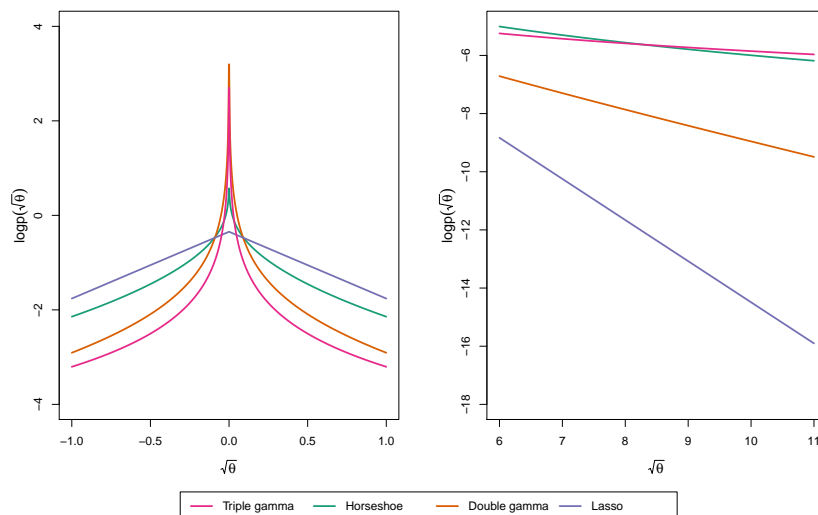

Figure 1. Marginal prior distribution of $\sqrt{\theta_j}$ under the triple gamma prior with $a^{\xi} = c^{\xi} = 0.1$ with $\kappa_B^2 = 2$, in comparison to the Horseshoe prior with $\phi^{\xi} = 1$, the double gamma prior with $a^{\xi} = 0.1$ and $\kappa_B^2 = 2$ and the Lasso prior with $\kappa_B^2 = 2$. Spike (left-hand side) and tail (right-hand side) of the marginal prior.

Lemma 1. *For $a^{\tilde{\xi}} > 0$ and $c^{\tilde{\xi}} > 0$, the triple gamma prior (4) has the following alternative representations:*

$$(a) \quad \sqrt{\theta}_j | \tilde{\xi}_j^2, c^{\tilde{\xi}}, \kappa_B^2 \sim t_{2c^{\tilde{\xi}}} \left(0, \frac{2}{\kappa_B^2} \tilde{\xi}_j^2 \right), \quad \tilde{\xi}_j^2 | a^{\tilde{\xi}} \sim \mathcal{G} \left(a^{\tilde{\xi}}, a^{\tilde{\xi}} \right), \tag{10}$$

$$(b) \quad \sqrt{\theta}_j | \check{\xi}_j^2, c^{\tilde{\xi}}, \kappa_B^2 \sim t_{2c^{\tilde{\xi}}} \left(0, \frac{2}{a^{\tilde{\xi}} \kappa_B^2} \check{\xi}_j^2 \right), \quad \check{\xi}_j^2 | a^{\tilde{\xi}} \sim \mathcal{G} \left(a^{\tilde{\xi}}, 1 \right). \tag{11}$$

Additional representations for $0 < a^{\tilde{\xi}} < \infty$ and $0 < c^{\tilde{\xi}} < \infty$ based on $\phi^{\tilde{\xi}} = \frac{2c^{\tilde{\xi}}}{\kappa_B^2 a^{\tilde{\xi}}}$ are

$$(c) \quad \sqrt{\theta}_j | \check{\xi}_j^2, \check{\kappa}_j^2, \phi^{\tilde{\xi}} \sim \mathcal{N} \left(0, \phi^{\tilde{\xi}} \check{\xi}_j^2 / \check{\kappa}_j^2 \right), \quad \check{\xi}_j^2 | a^{\tilde{\xi}} \sim \mathcal{G} \left(a^{\tilde{\xi}}, 1 \right), \quad \check{\kappa}_j^2 | c^{\tilde{\xi}} \sim \mathcal{G} \left(c^{\tilde{\xi}}, 1 \right), \tag{12}$$

$$(d) \quad \sqrt{\theta}_j | \tilde{\psi}_j^2, \phi^{\tilde{\xi}} \sim \mathcal{N} \left(0, \phi^{\tilde{\xi}} \tilde{\psi}_j^2 \right), \quad \tilde{\psi}_j^2 | a^{\tilde{\xi}}, c^{\tilde{\xi}} \sim \mathcal{BP} \left(a^{\tilde{\xi}}, c^{\tilde{\xi}} \right), \tag{13}$$

where $\mathcal{BP} \left(a^{\tilde{\xi}}, c^{\tilde{\xi}} \right)$ is the beta-prime distribution.[2]

2.3. Relation of the Triple Gamma to Other Shrinkage Priors

The triple gamma prior can be related to the very active research on shrinkage priors in a Bayesian framework in various ways. On the one hand, popular priors for variance parameters introduced as robust alternatives to the inverse gamma prior are special cases of the triple gamma, see Table 1. For instance, in (8), ψ_j^2 converges a.s. to 1, as $a^{\tilde{\xi}} \to \infty$ and $c^{\tilde{\xi}} \to \infty$, and the triple gamma reduces to a normal distribution for $\sqrt{\theta}_j$, applied for univariate TVP models (Frühwirth-Schnatter 2004) and unobserved component state space model (Frühwirth-Schnatter and Wagner 2010). For $c^{\tilde{\xi}} \to \infty$, $F \left(2a^{\tilde{\xi}}, 2c^{\tilde{\xi}} \right)$ converges to the $\mathcal{G} \left(a^{\tilde{\xi}}, a^{\tilde{\xi}} \right)$ distribution and the triple gamma reduces to the Bayesian Lasso for $a^{\tilde{\xi}} = 1$ (Belmonte et al. 2014) and otherwise to the double gamma (Bitto and Frühwirth-Schnatter 2019) applied in sparse TVP models.

Gelman (2006) introduced the half-t and the half-Cauchy prior for variance parameters in hierarchical models, by assuming that $\sqrt{\theta}_j$ follows a "folded" t-distribution, that is, a t-distribution truncated to $[0, \infty)$, see also Polson and Scott (2012b). In (10), $\tilde{\xi}_j^2$ converges a.s. to 1 as $a^{\tilde{\xi}} \to \infty$ and the triple gamma reduces to a $t_{2c^{\tilde{\xi}}}$- distribution and to the Cauchy distribution for $c^{\tilde{\xi}} = 1/2$, however without being "folded", since we allow $\sqrt{\theta}_j$ to take on negative values, as explained in Section 2.1.

---

[2]　Note that the $X \sim \mathcal{BP}(a, b)$-distribution has pdf

$$f(x) = \frac{1}{B(a, b)} \frac{x^{a-1}}{(1+x)^{a+b}}.$$

Furthermore, $Y = X / (1 + X)$ follows the $\mathcal{B}(a, b)$-distribution.

Table 1. Priors on $\sqrt{\theta}_j$ which are equivalent to (top) or special cases of (bottom) the triple gamma prior.

Prior for $\sqrt{\theta}_j$		a^ξ	c^ξ	κ_B^2	ϕ^ξ
$\mathcal{N}\left(0,\psi_j^2\right),\psi_j^2 \sim \mathcal{GG}\left(a^\xi,c^\xi,\phi^\xi\right)$	normal-gamma-gamma	a^ξ	c^ξ	$\frac{2c^\xi}{\phi^\xi a^\xi}$	ϕ^ξ
$\mathcal{N}\left(0,\frac{1}{\kappa_j}-1\right),\kappa_j \sim \mathcal{TPB}\left(a^\xi,c^\xi,\phi^\xi\right)$	generalized beta mixture	a^ξ	c^ξ	$\frac{2c^\xi}{\phi^\xi a^\xi}$	ϕ^ξ
$\mathcal{N}\left(0,\psi_j^2\right),\psi_j^2 \sim \text{SBeta2}\left(a^\xi,c^\xi,\phi^\xi\right)$	hierarchical scaled beta2	a^ξ	c^ξ	$\frac{2c^\xi}{\phi^\xi a^\xi}$	ϕ^ξ
$\mathcal{DE}\left(0,\sqrt{2}\,\psi_j\right),\psi_j^2 \sim \mathcal{G}\left(c^\xi,\frac{1}{\lambda^2}\right)$	normal-exponential-gamma	1	c^ξ	$2\lambda^2 c^\xi$	$\frac{1}{\lambda^2}$
$\mathcal{N}\left(0,\tau^2\psi_j^2\right),\psi_j \sim t_1$	Horseshoe	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{\tau^2}$	τ^2
$\mathcal{N}\left(0,\frac{1}{\kappa_j}-1\right),\kappa_j \sim \mathcal{B}\left(1/2,1\right)$	Strawderman-Berger	$\frac{1}{2}$	1	4	1
$\mathcal{N}\left(0,\tau^2\tilde{\xi}_j\right),\tilde{\xi}_j \sim \mathcal{G}\left(a^\xi,a^\xi\right)$	double gamma	a^ξ	∞	$\frac{2}{\tau^2}$	-
$\mathcal{N}\left(0,\tau^2\tilde{\xi}_j\right),\tilde{\xi}_j \sim \mathcal{E}\left(1\right)$	Lasso	1	∞	$\frac{2}{\tau^2}$	-
$t_\nu\left(0,\tau^2\right)$	half-t	∞	$\frac{\nu}{2}$	$\frac{2}{\tau^2}$	-
$t_1\left(0,\tau^2\right)$	half-Cauchy	∞	$\frac{1}{2}$	$\frac{2}{\tau^2}$	-
$\mathcal{N}\left(0,B_0\right)$	normal	∞	∞	$\frac{2}{B_0}$	-

On the other hand, the triple gamma prior is related to popular shrinkage priors in regression models. It extends the generalized beta mixture prior introduced by Armagan et al. (2011) for variable selection in regression models,

$$\beta_j|\xi_j^2 \sim \mathcal{N}\left(0,\xi_j^2\right), \quad \xi_j^2 \sim \mathcal{G}\left(a^\xi,\lambda_j\right), \quad \lambda_j \sim \mathcal{G}\left(c^\xi,\phi^\xi\right),$$

to variance selection in state space and TVP models. This is evident from rewriting (4) as $\xi_j^2 \sim \mathcal{G}\left(a^\xi,\lambda_j\right), \lambda_j \sim \mathcal{G}\left(c^\xi,\phi^\xi\right)$. We exploit this relationship in Section 3.1 to investigate the shrinkage profile of a triple gamma prior. Using Armagan et al. (2011, Definition 2), the triple gamma prior can be written as

$$\sqrt{\theta}_j|\rho_j \sim \mathcal{N}\left(0,1/\rho_j-1\right), \quad \rho_j|a^\xi,c^\xi,\phi^\xi \sim \mathcal{TPB}\left(a^\xi,c^\xi,\phi^\xi\right), \tag{14}$$

where $\mathcal{TPB}\left(a^\xi,c^\xi,\phi^\xi\right)$ is the three-parameter beta distribution with density:

$$p(\rho_j) = \frac{1}{B(a^\xi,c^\xi)}(\phi^\xi)^{c^\xi}\rho_j^{c^\xi-1}(1-\rho_j)^{a^\xi-1}\left(1+(\phi^\xi-1)\rho_j\right)^{-(a^\xi+c^\xi)}. \tag{15}$$

From (14) and (15), it becomes evident that the Strawderman-Berger prior $\sqrt{\theta}_j|\rho_j \sim \mathcal{N}\left(0,1/\rho_j-1\right), \rho_j \sim \mathcal{B}\left(1/2,1\right)$ (Berger 1980; Strawderman 1971) is that special case of the triple gamma prior where $\phi^\xi = 1$, $a^\xi = 1/2$, and $c^\xi = 1$.

The special case of a triple gamma, where $a^\xi = c^\xi = 1/2$, corresponds to a Horseshoe prior (Carvalho et al. 2009 2010) on $\sqrt{\theta}_j$ with global shrinkage parameter $\tau^2 = 2/\kappa_B^2$, since $\psi_j^2 \sim F\left(1,1\right)$ implies that $\psi_j \sim t_1$. The Horseshoe prior has been introduced for variable selection in regression models and has been shown to have excellent theoretical properties in this context for the "nearly black" case (van der Pas et al. 2014). The triple gamma is a generalization of the Horseshoe prior, with a similar shrinkage profile, however with much more mass close to the corner solutions. Most importantly, as will be discussed in Section 3.1, this leads to a BMA-type behaviour of the triple gamma prior for small values of a^ξ and c^ξ.

The vast literature on shrinkage priors contains many more related priors. Rescaling $\tilde{\xi}_j^2 = 2/(\kappa_B^2)\psi_j^2$ in (8), for instance, yields a representation involving a scaled beta2 distribution,[3]

$$\sqrt{\theta_j}|\tilde{\xi}_j^2 \sim \mathcal{N}\left(0,\tilde{\xi}_j^2\right), \qquad \tilde{\xi}_j^2|a^\xi,c^\xi,\phi^\xi \sim \text{SBeta2}\left(a^\xi,c^\xi,\phi^\xi\right), \tag{16}$$

as is easily derived from (A2). The scaled beta2 was introduced by Pérez et al. (2017) in hierarchical models as a robust prior for scale parameters, $\sqrt{\theta_j}$, and variance parameters, θ_j, alike. Based on (16), the triple gamma can be seen as a hierarchical extension of this prior which puts a scaled beta2 distribution on the scaling parameter $\tilde{\xi}_j^2$ of a Gaussian prior for $\sqrt{\theta_j}$, see Table 1. Griffin and Brown (2017) termed prior (16) gamma-gamma distribution, denoted by GG $\left(a^\xi,c^\xi,\phi\right)$.

For $a^\xi = 1$, the triple gamma reduces to the normal-exponential-gamma which has a representation as a scale-mixture of double exponential $\mathcal{DE}\left(0,\sqrt{2}\psi_j\right)$-distributions, see Table 1.

It has been considered for variable selection in regression models (Griffin and Brown 2011) and locally adaptive B-spline models (Scheipl and Kneib 2009). The R2-D2 prior suggested by Zhang et al. (2017) for high-dimensional regression models is another special case of the triple gamma. It reads

$$\beta_j \sim \mathcal{N}\left(0,\sigma^2\phi_j\omega\right), \quad (\phi_1,\ldots,\phi_d) \sim \mathcal{D}\left(a^\tau,\ldots,a^\tau\right), \quad \omega \sim \mathcal{G}\left(a,\tau\right), \quad \tau \sim \mathcal{G}\left(b,1\right),$$

where $a = da^\tau$ and σ^2 is the residual error variance of the regression model. As shown by Zhang et al. (2017), this implies the following prior for the coefficient of determination: $R^2 \sim \mathcal{B}\left(a,b\right)$ which motivates holding a fixed, while a^τ decreases as d increases. Using that $\phi_j\omega \sim \mathcal{G}\left(a^\tau,\tau\right)$, we can show that the R2-D2 prior is equivalent to the following hierarchical normal gamma prior applied in Bitto and Frühwirth-Schnatter (2019) for TVP models:

$$\beta_j|\tau_j^2 \sim \mathcal{N}\left(0,\tau_j^2\right), \quad \tau_j^2 \sim \mathcal{G}\left(a^\tau,a^\tau\lambda_B^2/2\right), \quad \lambda_B^2 \sim \mathcal{G}\left(b,2\sigma^2/a^\tau\right).$$

The popular Dirichlet-Laplace prior, $\sqrt{\theta_j}|\psi_j \sim \mathcal{DE}\left(0,\psi_j\right)$, however, is not related to the triple gamma as the prior *scale* ψ_j rather than the prior *variance* ψ_j^2 follows a gamma distribution, see again Table 1.

2.4. Using the Triple Gamma for Variance Selection in TVP Models

A challenging question is how to choose the parameters a^ξ, c^ξ and κ_B^2 or ϕ^ξ of the triple gamma prior in the context of variance selection for TVP models. In addition, in a TVP context, the shrinkage parameters a^τ, c^τ and λ_B^2 or $\phi^\tau = 2c^\tau/(a^\tau\lambda_B^2)$ for the prior (7) of the initial values β_j have to be selected.

In high-dimensional settings it is appealing to have a prior that addresses two major issues: first, high concentration around the origin to favor strong shrinkage of small variances toward zero; second, heavy tails to introduce robustness to large variances and to avoid over-shrinkage. For the triple gamma prior, both issues are addressed through the choice of a^ξ and c^ξ, see Theorem 2. First of all, we need values $0 < a^\xi \leq 0.5$ to induce a pole at 0. Second, values of $0 < c^\xi < 0.5$ will lead to very heavy tails. For very small values of a^ξ and c^ξ, the triple Gamma is a proper prior that behaves nearly as the improper normal-Jeffrey's prior (Figueiredo 2003), where $p(\sqrt{\theta_j}) \propto 1/\sqrt{\theta_j}$ and $p(\rho_j) \propto \rho_j^{-1}(1-\rho_j)^{-1}$.

Ideally, we would place a hyper prior distribution on all shrinkage parameters which would allow us to learn the global and the local degree of sparsity, both for the variances and the initial values. Such a

---

3　The pdf of a SBeta2 (a,c,ϕ)-distribution reads:

$$f(x) = \frac{1}{\phi^a B(a,c)}x^{a-1}(1+x/\phi)^{-(a+c)},$$

hierarchical triple gamma prior introduces dependence among the local shrinkage parameters $\xi_1^2, \ldots, \xi_d^2$ in (4) and, consequently, among $\theta_1, \ldots, \theta_d$ in the joint (marginal) prior $p(\theta_1, \ldots, \theta_d)$. Introducing such dependence is desirable in that it allows to learn the degree of variance sparsity in TVP models, meaning that how much a variance is shrunken toward zero depends on how close the other variances are to zero. However, first naïve approaches with rather uninformative, independent priors on κ_B^2, a^ξ, c^ξ and λ_B^2, a^τ, c^τ were not met with much success and we found it necessary to carefully design appropriate hyper priors.

Hierarchical versions of the Bayesian Lasso (Belmonte et al. 2014) and the double gamma prior (Bitto and Frühwirth-Schnatter 2019) in TVP models are based on the gamma prior $\kappa_B^2 \sim \mathcal{G}(d_1, d_2)$. Interestingly, this choice can be seen as a heavy-tailed extension of both priors, where each marginal density $p(\sqrt{\theta_j}|d_1, d_2)$ follows a triple gamma prior with the same parameter a^ξ (being equal to one for the Bayesian Lasso) and tail index $c^\xi = d_1$. In light of this relationship, it is not surprising that very small values of d_1 were applied in these papers to ensure heavy tails of $p(\sqrt{\theta_j}|d_1, d_2)$. Since a triple gamma prior already has heavy tails, we choose a different hyperprior in the present paper.

For the case $a^\xi = c^\xi = 1/2$, the global shrinkage parameter τ of the Horseshoe prior typically follows a Cauchy prior, $\tau \sim t_1$ (Bhadra et al. 2017b; Carvalho et al. 2009), see also Bhadra et al. (2019, Section 5). The relationship $\phi^\xi = 2/\kappa_B^2 = \tau^2$ between the various global shrinkage parameters (see Table 1) implies in this case $\phi^\xi \sim F(1, 1)$ or, equivalently, $\kappa_B^2/2 \sim F(1, 1)$.

For a triple gamma prior with arbitrary a^ξ and c^ξ, this is a special case of the following prior:

$$\frac{\kappa_B^2}{2} \bigg| a^\xi, c^\xi \sim F\left(2a^\xi, 2c^\xi\right), \tag{17}$$

which will be motivated in Section 3.2. Under this prior, the triple gamma prior exhibits BMA-like behavior with a uniform prior on an appropriately defined model size (see Theorem 3). Prior (17) is equivalent to following representations:

$$\kappa_B^2|a^\xi \sim \mathcal{G}\left(a^\xi, d_2\right), \qquad d_2|a^\xi, c^\xi \sim \mathcal{G}\left(c^\xi, \frac{2c^\xi}{a^\xi}\right), \tag{18}$$
$$\phi^\xi|a^\xi, c^\xi \sim \mathcal{BP}\left(c^\xi, a^\xi\right).$$

Concerning a^ξ and c^ξ, we choose the following priors:

$$2a^\xi \sim \mathcal{B}(\alpha_{a^\xi}, \beta_{a^\xi}), \qquad 2c^\xi \sim \mathcal{B}(\alpha_{c^\xi}, \beta_{c^\xi}). \tag{19}$$

Hence, we are restricting the support of a^ξ and c^ξ to $(0, 0.5)$, following the insights brought to us by Theorem 2.

We follow a similar strategy for the parameters a^τ, c^τ and λ_B^2 (ϕ^τ) of the prior (7) of the initial values β_j:

$$\frac{\lambda_B^2}{2} \bigg| a^\tau, c^\tau \sim F(2a^\tau, 2c^\tau), \quad 2a^\tau \sim \mathcal{B}(\alpha_{a^\tau}, \beta_{a^\tau}), \quad 2c^\tau \sim \mathcal{B}(\alpha_{c^\tau}, \beta_{c^\tau}), \tag{20}$$

which is equivalent to $\lambda_B^2|a^\tau \sim \mathcal{G}(a^\tau, e_2)$, $e_2|a^\tau, c^\tau \sim \mathcal{G}(c^\tau, 2c^\tau/a^\tau)$, and $\phi^\tau|a^\tau, c^\tau \sim \mathcal{BP}(c^\tau, a^\tau)$.

An interesting special case is the "symmetric" triple gamma, where $a^\xi = c^\xi$. Despite this constraint, the favourable shrinkage behaviour is preserved and decreasing $a^\xi = c^\xi$ toward zero simultaneously leads to a high concentration around the origin and a heavy-tailed behaviour. For a symmetric triple gamma prior, the global shrinkage parameter ϕ^ξ is independent of a^ξ and c^ξ and is related to the global shrinkage parameters κ_B^2 through $\phi^\xi = 2/\kappa_B^2$. This induces shrinkage profiles that are symmetric around $1/2$, see Section 3.1. Interestingly, a symmetric triple gamma resolves the question whether to choose a gamma or an inverse gamma prior for a variance parameter ψ_j^2. It implies

the same symmetric beta-prime distribution on the variance, $\psi_j^2 \sim \mathrm{F}\left(2a^\xi, 2a^\xi\right) = \mathcal{BP}\left(a^\xi, a^\xi\right)$, and the information, $(\psi_j^2)^{-1} \sim \mathcal{BP}\left(a^\xi, a^\xi\right)$, and can be represented as a gamma prior with the scale arising from an inverse gamma prior or, equivalently, as an inverse gamma prior with the scale arising from a gamma prior:

$$\psi_j^2 = \breve{\xi}_j^2 \times \frac{1}{\breve{\kappa}_j^2}, \quad (\psi_j^2)^{-1} = \breve{\kappa}_j^2 \times \frac{1}{\breve{\xi}_j^2}, \qquad \breve{\xi}_j^2 \sim \mathcal{G}\left(a^\xi, 1\right), \ \breve{\kappa}_j^2 \sim \mathcal{G}\left(a^\xi, 1\right).$$

3. Shrinkage Profiles and BMA-Like Behavior

3.1. Shrinkage Profiles

In the sparse normal-means problem where $\mathbf{y}|\boldsymbol{\beta} \sim \mathcal{N}_d\left(\boldsymbol{\beta}, \sigma^2 \mathbf{I}_d\right)$ and $\sigma^2 = 1$, the parameter $\rho_j = 1/(1 + \psi_j^2)$ appearing in (14) is known as shrinkage factor and plays a fundamental role for comparing different shrinkage priors, as ρ_j determines shrinkage toward 0.

Also in a variance selection context, it is evident from (14) that values of $\rho_j \approx 0$ will introduce no shrinkage on θ_j, whereas values of $\rho_j \approx 1$ will introduce strong shrinkage of θ_j toward 0. Hence, the prior $p(\rho_j)$, also called shrinkage profile, will play an instrumental role in the behaviour of different shrinkage priors. Following Carvalho et al. (2010), shrinkage priors are often compared in terms of the prior they imply on ρ_j, that is, how they handle shrinkage for small "observations" (in our case innovations) and how robust they are to large "observations". Note that we ideally want a shrinkage profile that has a pole in zero (heavy tails to avoid over-shrinking signals) and a pole in one (spikiness to shrink noise). The Horseshoe prior, for example, implies $\rho_j \sim \mathcal{B}\left(1/2, 1/2\right)$ which is a shrinkage profile that takes this much desired form of a "horseshoe", see Figure 2.

For the triple gamma prior, the shrinkage profile is given by the three-parameter beta prior $p(\rho_j)$ provided in (15). For $\phi^\xi = 1$, $\rho_j \sim \mathcal{B}\left(c^\xi, a^\xi\right)$ and $\kappa_B^2 = 2c^\xi/a^\xi$. Choosing small values $a^\xi \ll 1$ will put prior mass close to 1, choosing small values $c^\xi \ll 1$ will put prior mass close to 0, whereas values for both a^ξ and c^ξ smaller than one will induce the form of a horseshoe prior for ρ_j. Evidently, for $\phi^\xi = 1$, a symmetric triple gamma prior with $a^\xi = c^\xi$ implies a Horseshoe prior for ρ_j that is symmetric around 0.5. This is illustrated in Figure 2 for a symmetric triple gamma with $a^\xi = c^\xi = 0.1$.
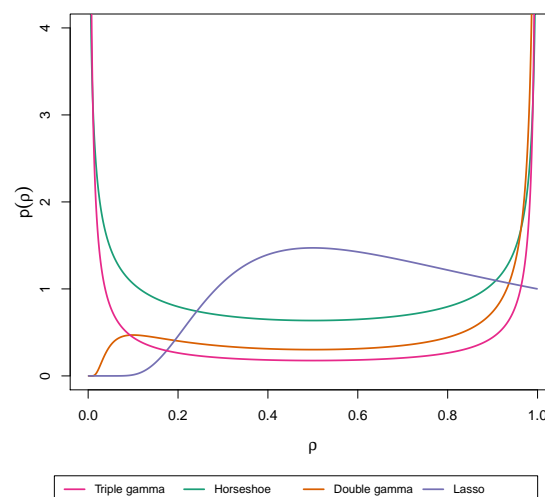

Figure 2. Marginal univariate shrinkage profile under the triple gamma prior with $a^\xi = c^\xi = 0.1$, in comparison to the Horseshoe prior, the double gamma prior with $a^\xi = 0.1$ and the Lasso prior. $\kappa_B^2 = 2$ for all the prior specifications.

In Figure 2 we can also see the shrinkage profile for the Bayesian Lasso and the double gamma, which correspond to a triple gamma where $c^{\tilde{\xi}} \to \infty$.[4] For the Bayesian Lasso with $a^{\tilde{\xi}} = 1$, it is clear that the shrinkage profile $p(\rho_j)$ converges to a constant for $\rho_j \to 1$, while there is no mass around $\rho_j = 0$. This means that this prior tends to over-shrink signals, while not shrinking the noise completely to zero. A double gamma prior with $a^{\tilde{\xi}} < 1$ has the potential to shrink the noise completely to zero, as $p(\rho_j)$ has a pole at $\rho_j = 1$, but $p(\rho_j)$ has also zero mass around $\rho_j = 0$, meaning the prior encourages over-shrinking of signals.

When we make κ_B^2 random, we obtain a "prior density" of shrinkage profiles, see Figure 3. We can see that such hierarchical versions of the Lasso and the double gamma have shrinkage profiles that resemble the ones of the Horseshoe and the triple gamma. We have used $\kappa_B^2 \sim \mathcal{G}(0.01, 0.01)$ for the Lasso and the double gamma, $2/\kappa_B^2 \sim F(1, 1)$ for the Horseshoe and $2/\kappa_B^2 \sim F(0.2, 0.2)$ for the triple gamma, see Section 2.4.
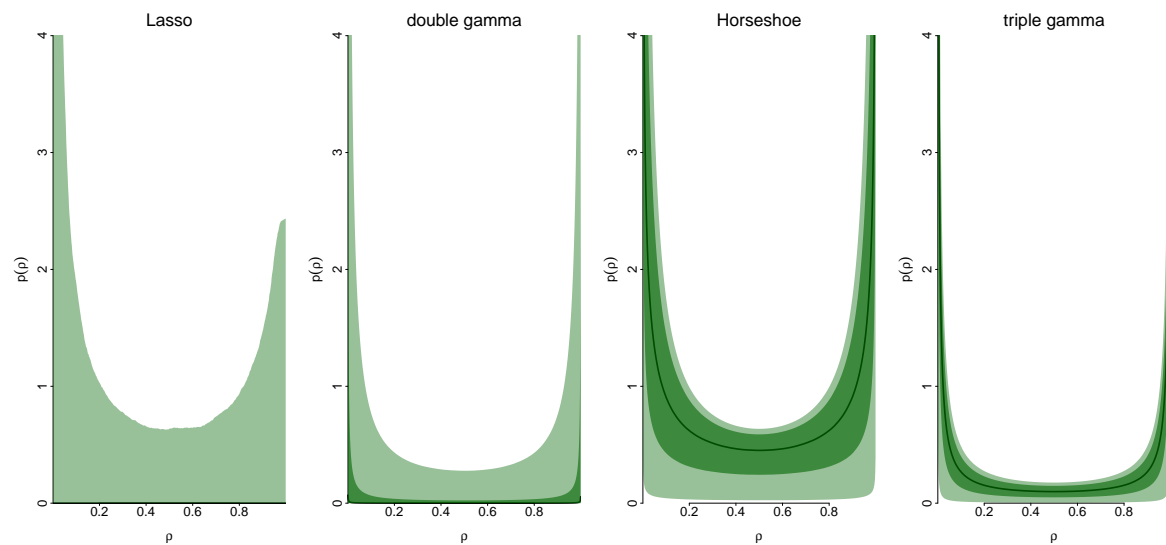

Figure 3. "Prior density" of shrinkage profiles for (from left to right) a Lasso prior, a double gamma prior with $a^{\tilde{\xi}} = 0.2$, a Horseshoe prior and a triple gamma prior with $a^{\tilde{\xi}} = c^{\tilde{\xi}} = 0.1$, when κ_B^2 is random. The solid line is the median, while the shaded areas represent 50% and 95 % prior credible bands. We have used $\kappa_B^2 \sim \mathcal{G}(0.01, 0.01)$ for the Lasso and the double gamma, $2/\kappa_B^2 \sim F(1, 1)$ for the Horseshoe and $2/\kappa_B^2 \sim F(0.2, 0.2)$ for the triple gamma.

3.2. BMA-Type Behaviour

Bayesian model averaging (BMA) provides to statisticians and practitioners an essential and coherent tool to account for model uncertainty. In a multiple regression setting, the uncertainty is inherent in the choice of variables to be included. In a TVP framework, there is additional uncertainty about the time-variation of the state parameters, that is, which explanatory variables have a static and which ones a dynamic effect on the response variable. In this section, we show that the triple gamma prior mimics the typical BMA behavior, thus allowing us to incorporate model uncertainty with respect to time variation.

---

[4]　Using (3), we obtain the following prior for $\rho_j = 1/(1 + \psi_j^2)$ by the law of transformation of densities:

$$p(\rho_j) = \frac{1}{\Gamma(a^{\tilde{\xi}})} \left(\frac{a^{\tilde{\xi}} \kappa_B^2}{2} \right)^{a^{\tilde{\xi}}} (1 - \rho_j)^{a^{\tilde{\xi}} - 1} \rho_j^{-(a^{\tilde{\xi}} + 1)} \exp \left(- \left(\frac{1 - \rho_j}{\rho_j} \right) \frac{a^{\tilde{\xi}} \kappa_B^2}{2} \right).$$

From the perspective of Bayesian model averaging, an ideal approach for handling sparsity in TVP models would be the use of discrete mixture priors as suggested in Frühwirth-Schnatter and Wagner (2010),

$$p(\sqrt{\theta_j}) = (1 - \pi)\delta_0 + \pi \cdot p_{\text{slab}}(\sqrt{\theta_j}), \tag{21}$$

with δ_0 being a Dirac measure at 0, while $p_{\text{slab}}(\sqrt{\theta_j})$ is the prior for non-zero variances. In terms of shrinkage profiles, the discrete mixture prior (21) has a spike at $\rho_j = 1$, with probability $1 - \pi$, and a lot of prior mass at $\rho_j = 0$, provided that the tails of $p_{\text{slab}}(\sqrt{\theta_j})$ are heavy enough. The mixture prior (21) is considered the "gold standard" in BMA, both theoretically and empirically, see for example, Johnstone and Silverman (2004). However, MCMC inference under this prior is extremely challenging. As opposed to this, MCMC inference for the triple gamma prior is straightforward, see Section 4.

In this section, we relate the triple gamma prior to BMA based on the discrete mixture prior (21). An interesting insight is that the triple gamma prior shows a behaviour very similar to a discrete mixture prior, if both a^ξ and c^ξ approach zero. This induces BMA-type behaviour on the joint shrinkage profile $p(\rho_1, \ldots, \rho_d)$, with a spike at all corner solutions, where some ρ_j are very close to one, whereas the remaining ones are very close to zero.

The bivariate shrinkage profiles shown in Figure 4 give us some intuition about the convergence of a symmetric triple gamma prior with $a^\xi = c^\xi \to 0$ toward a discrete spike and slab mixture. As opposed to the Lasso and the double gamma prior, the Horseshoe and the triple gamma prior put nearly all prior mass on the "corner solutions", which correspond to the four possibilities (a) $\rho_1 = \rho_2 = 0$, that is, no shrinkage on θ_1 and θ_2, (b) $\rho_1 = 1, \rho_2 = 0$, that is, shrinkage of θ_1 toward 0 and no shrinkage on θ_2, (c) $\rho_1 = 0, \rho_2 = 1$, that is, shrinkage of θ_1 toward 0 and no shrinkage on θ_2, and (d) $\rho_1 = \rho_2 = 1$, that is, shrinkage of both θ_1 and θ_2 toward 0.
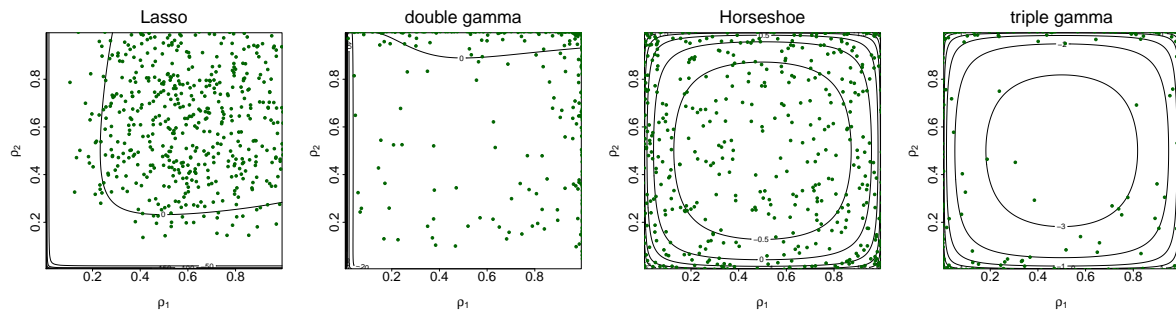

Figure 4. Bivariate shrinkage profile $p(\rho_1, \rho_2)$ for (from left to right) the Lasso prior, the double gamma prior with $a^\xi = 0.1$, the Horseshoe prior, and the triple gamma prior with $a^\xi = c^\xi = 0.1$, with $\kappa_B^2 = 2$ for all the priors. The contour plots of the bivariate shrinkage profile are shown, together with 500 samples from the bivariate prior distribution of the shrinkage parameters.

A very important aspect of BMA is that of choosing a prior for the model dimension, K, see for example, Fernández et al. (2001) and Ley and Steel (2009). In the discrete mixture prior (21), the distribution of K depends on the choice of π. Fixing π corresponds to a very informative prior on the model dimension, for example $\pi = 0.5$ assigns more prior probability to models of dimension $d/2$ and lower prior probability to empty or full models. In fact, let δ_j be the indicator that tells us if the j-th coefficient is included in the model, then we have that $K = \sum_{j=1}^d \delta_j \sim \text{Binom}(d, \pi)$. Placing a uniform prior on π has been shown to be a good choice, since it corresponds to placing a prior on K which is uniform on $\{0, \ldots, d\}$. Note that π will be learned using information from all the variables. In this sense, π is a global shrinkage parameter which will adapt to the degree of sparsity.

Following ideas in Carvalho et al. (2009), we believe that a natural way to perform variable selection in the continuous shrinkage prior framework is though thresholding. Specifically, we say that when $(1 - \rho_j) > 0.5$, or $\rho_j < 0.5$, the variable is included, otherwise it is not. Notice that this

classification via thresholding makes perfect sense in the case of a triple gamma of which the Horseshoe is a special case, but less so for a Lasso or double gamma prior, even if the shrinkage profile shows a Horseshoe-like behaviour for hierarchical versions of these priors (see again Figure 3). Notice that thresholding implies a prior on the model dimension K. Specifically,

$$K = \sum_{j=1}^{d} I\{\rho_j < 0.5\} \sim \text{Binom}\left(d, \pi^{\xi}\right), \quad \pi^{\xi} = \Pr(\rho_j < 0.5),$$ (22)

where $\rho_j | a^{\xi}, \phi^{\xi} \sim \mathcal{TPB}\left(a^{\xi}, b^{\xi}, \phi^{\xi}\right)$, see (15). The choice of ϕ^{ξ} (or κ_B^2) will strongly impact the prior on K. For a symmetric triple gamma with $a^{\xi} = c^{\xi}$, for instance, and fixed $\phi^{\xi} = 1$, that is $\kappa_B^2 = 2$, we obtain $K \sim \text{Binom}(d, 0.5)$, since $\pi^{\xi} = 0.5$ regardless of a^{ξ}. Hence, we have to face similar problems as with fixing $\pi = 0.5$ for the discrete mixture prior (21).

Placing a hyper prior on ϕ^{τ} and ϕ^{ξ} (or equivalent ones on λ_B^2 and κ_B^2), as we did in Section 2.4, is as vital for BMA-type variable and variance selection through the triple gamma prior, as making π random is for the discrete mixture prior (21). Ideally, we would like to have a uniform distribution on the model size K. We show in Theorem 3 that the hyperprior for κ_B^2 defined in (17) achieves exactly this goal, since π^{ξ} is uniformly distributed, see Appendix A for a proof.

Theorem 3. *For a hierarchical triple gamma prior with fixed $a^{\xi} > 0$ and $c^{\xi} > 0$ the probability π^{ξ} defined in (22) follows a uniform distribution, $\pi^{\xi} \sim \mathcal{U}[0, 1]$, under the hyper prior*

$$\frac{\kappa_B^2}{2} \bigg| a^{\xi}, c^{\xi} \sim \text{F}\left(2a^{\xi}, 2c^{\xi}\right),$$ (23)

or, equivalently, under the hyper prior

$$\phi^{\xi} | a^{\xi}, c^{\xi} \sim \mathcal{BP}\left(c^{\xi}, a^{\xi}\right).$$ (24)

Finally, it is important to point out that the thresholding approach allows us to estimate posterior inclusion probabilities, that is the probability that the corresponding variable is included in the model or, in the case of variance selection, that the corresponding parameter is time varying. In our simulations (Section 5.3) and in our application (Section 5.4), we will estimate the posterior inclusion probabilities obtained under different shrinkage priors.

4. MCMC Algorithm

Let $\boldsymbol{y} = (y_1, \dots, y_T)$ be the vector of time series observations and let $\boldsymbol{z}$ be the set of all latent variables and unknown model parameters in a TVP model. Moreover, let $\boldsymbol{z}_{-x}$ denote the set of all unknowns but x. Bayesian inference based on MCMC sampling from the posterior $p(\boldsymbol{z}|\boldsymbol{y})$ is summarized in Algorithm 1. The hierarchical priors introduced in Section 2.4 are employed, where κ_B^2 follows (17), (a^{ξ}, c^{ξ}) follow (19), and $(a^{\tau}, c^{\tau}, \lambda_B^2)$ follow (20). For certain sampling steps, the hierarchical representation (18) is used for κ_B^2, and similarly for λ_B^2.

Algorithm 1 extends several existing algorithms such as the MCMC schemes introduced for the Horseshoe prior by Makalic and Schmidt (2016) and for the double gamma prior by Bitto and Frühwirth-Schnatter (2019). We exploit various representations of the triple gamma prior given in Lemma 1 and choose representation (12) as the baseline representation of our MCMC algorithm:

$$\beta_j | \check{\tau}_j^2, \check{\lambda}_j^2, \phi^{\tau} \sim \mathcal{N}\left(0, \phi^{\tau} \check{\tau}_j^2 / \check{\lambda}_j^2\right), \quad \check{\tau}_j^2 | a^{\tau} \sim \mathcal{G}\left(a^{\tau}, 1\right), \quad \check{\lambda}_j^2 | c^{\tau} \sim \mathcal{G}\left(c^{\tau}, 1\right),$$

$$\sqrt{\theta_j} | \check{\xi}_j^2, \check{\kappa}_j^2, \phi^{\xi} \sim \mathcal{N}\left(0, \phi^{\xi} \check{\xi}_j^2 / \check{\kappa}_j^2\right), \quad \check{\xi}_j^2 | a^{\xi} \sim \mathcal{G}\left(a^{\xi}, 1\right), \quad \check{\kappa}_j^2 | c^{\xi} \sim \mathcal{G}\left(c^{\xi}, 1\right),$$

where $\phi^\tau = 2c^\tau/(\lambda_B^2 a^\tau)$ and $\phi^\xi = 2c^\xi/(\kappa_B^2 a^\xi)$. All conditional distributions in our MCMC scheme are available in closed form, except for the ones for a^ξ, c^ξ, a^τ and c^τ, for which we will resort to a Metropolis-Hastings (MH) step within Gibbs. Several conditional distributions are the same as for the double gamma prior and we apply Algorithm 1 of Bitto and Frühwirth-Schnatter (2019). We provide more details on the derivation of the various densities in Appendix B.

Algorithm 1. MCMC inference for TVP models under the triple gamma prior.

Choose starting values for all global shrinkage parameters $(a^\tau, c^\tau, \lambda_B^2, a^\xi, c^\xi, \kappa_B^2)$ and local shrinkage parameters $\{\check{\tau}_j^2, \check{\lambda}_j^2, \check{\xi}_j^2, \check{\kappa}_j^2\}_{j=1}^d$, and repeat the following steps:

(a) *Define for $j = 1, \ldots, d$, $\tau_j^2 = \phi^\tau \check{\tau}_j^2 / \check{\lambda}_j^2$ and $\xi_j^2 = \phi^\xi \check{\xi}_j^2 / \check{\kappa}_j^2$ and sample from the posterior $p(\tilde{\beta}_0, \ldots, \tilde{\beta}_T, \beta_1, \ldots, \beta_d, \sqrt{\theta_1}, \ldots, \sqrt{\theta_d} | \{\xi_j^2, \tau_j^2\}_{j=1}^d, y)$ using Algorithm 1, Steps (a), (b), and (c) in Bitto and Frühwirth-Schnatter (2019). In the homoscedastic case, use Step (f) of this algorithm to sample from $\sigma^2 | z_{-\sigma^2}, y$. For the SV model (2), sample the parameters μ, ϕ, and σ_η^2 as in Kastner and Frühwirth-Schnatter (2014), for example, using the R-package* `stochvol` *(Kastner 2016).*

(b) *Use the prior $p(\sqrt{\theta_j} | \check{\kappa}_j^2, a^\xi, c^\xi)$, marginalized w.r.t. $\check{\xi}_j^2$, to sample a^ξ from $p(a^\xi | z_{-a^\xi}, y)$ via a random walk MH step on $z = \log(a^\xi / (0.5 - a^\xi))$. Propose $a^{\xi,(*)} = 0.5 e^{z^*} / (1 + e^{z^*})$, where $z^* \sim \mathcal{N}\left(z^{(m-1)}, v^2\right)$ and $z^{(m-1)} = \log(a^{\xi,(m-1)} / (0.5 - a^{\xi,(m-1)}))$ depends on the previous value $a^{\xi,(m-1)}$ of a^ξ, accept $a^{\xi,(*)}$ with probability*

$$\min\left\{1, \frac{q_a(a^{\xi,(*)})}{q_a(a^{\xi,(m-1)})}\right\}, \qquad q_a(a^\xi) = p(a^\xi | z_{-a^\xi}, y) \, a^\xi (0.5 - a^\xi),$$

and update $\phi^\xi = 2c^\xi / (\kappa_B^2 a^\xi)$. Explicit forms for $p(a^\xi | z_{-a^\xi}, y)$ and $\log q_a(a^\xi)$ are provided in (A3) and (A4).

Similarly, use the prior $p(\beta_j | \check{\lambda}_j^2, a^\tau, c^\tau)$, marginalized w.r.t. to $\check{\tau}_j^2$, to sample a^τ via a random walk MH step and update $\phi^\tau = 2c^\tau / (a^\tau \lambda_B^2)$.

(c) *Sample $\check{\xi}_j^2$, $j = 1, \ldots, d$, from a generalized inverse Gaussian distribution, see (A5):*

$$\check{\xi}_j^2 | \check{\kappa}_j^2, \theta_j, a^\xi, \phi^\xi \sim \mathcal{GIG}\left(a^\xi - \frac{1}{2}, 2, \frac{\check{\kappa}_j^2 \theta_j}{\phi^\xi}\right). \tag{25}$$

Similarly, update $\check{\tau}_j^2$, $j = 1, \ldots, d$, conditional on a^τ:

$$\check{\tau}_j^2 | \beta_j, \check{\lambda}_j^2, a^\tau, \phi^\tau \sim \mathcal{GIG}\left(a^\tau - \frac{1}{2}, 2, \frac{\check{\lambda}_j^2 \beta_j^2}{\phi^\tau}\right).$$

(d) *Use the marginal Student-t distribution $p(\sqrt{\theta_j} | \check{\xi}_j^2, c^\xi, \kappa_B^2)$ given in (11) to sample c^ξ from $p(c^\xi | z_{-c^\xi}, y)$ via a random walk MH step on $z = \log(c^\xi / (0.5 - c^\xi))$. Propose $c^{\xi,(*)} = 0.5 e^{z^*} / (1 + e^{z^*})$, where $z^* \sim \mathcal{N}\left(z^{(m-1)}, v^2\right)$ and $z^{(m-1)} = \log(c^{\xi,(m-1)} / (0.5 - c^{\xi,(m-1)}))$ depends on the previous value $c^{\xi,(m-1)}$ of c^ξ, accept $c^{\xi,(*)}$ with probability*

$$\min\left\{1, \frac{q_c(c^{\xi,(*)})}{q_c(c^{\xi,(m-1)})}\right\}, \qquad q_c(c^\xi) = p(c^\xi | z_{-c^\xi}, y) \, c^\xi (0.5 - c^\xi),$$

and update $\phi^\xi = 2c^\xi / (\kappa_B^2 a^\xi)$. Explicit forms for $p(c^\xi | z_{-c^\xi}, y)$ and $\log q_c(c^\xi)$ are provided in (A6) and (A7).

Similarly, to sample c^τ via a random walk MH step use the marginal distribution of $\beta_j | \check{\tau}_j^2, a^\tau, c^\tau$ with respect to $\check{\lambda}_j^2$ and update $\phi^\tau = 2c^\tau / (a^\tau \lambda_B^2)$.

(e)   Sample $\check{\kappa}_j^2$, for $j = 1, \ldots, d$, from following gamma distribution, see (A8):

$$\check{\kappa}_j^2 | \theta_j, \check{\xi}_j^2, c^\xi, \phi^\xi \sim \mathcal{G}\left(\frac{1}{2} + c^\xi, \frac{\theta_j}{2\phi^\xi \check{\xi}_j^2} + 1\right). \tag{26}$$

Similarly, update $\check{\lambda}_j^2$, $j = 1, \ldots, d$, conditional on c^τ:

$$\check{\lambda}_j^2 | \beta_j, \check{\tau}_j^2, c^\tau, \phi^\tau \sim \mathcal{G}\left(\frac{1}{2} + c^\tau, \frac{\beta_j^2}{2\phi^\tau \check{\tau}_j^2} + 1\right).$$

(f)   Sample d_2 from $d_2 | a^\xi, c^\xi, \kappa_B^2 \sim \mathcal{G}\left(a^\xi + c^\xi, \kappa_B^2 + \frac{2c^\xi}{a^\xi}\right)$, see (A9); sample from κ_B^2 from following gamma distribution,

$$\kappa_B^2 | \{\theta_j, \check{\kappa}_j^2, \check{\xi}_j^2\}_{j=1}^d, a^\xi, c^\xi, d_2 \sim \mathcal{G}\left(\frac{d}{2} + a^\xi, \frac{a^\xi}{4c^\xi} \sum_{j=1}^d \frac{\check{\kappa}_j^2}{\check{\xi}_j^2} \theta_j + d_2\right), \tag{27}$$

see (A10), and update $\phi^\xi = 2c^\xi / (\kappa_B^2 a^\xi)$.

Similarly, sample e_2 from $e_2 | a^\tau, c^\tau, \lambda_B^2 \sim \mathcal{G}\left(a^\tau + c^\tau, \lambda_B^2 + \frac{2c^\tau}{a^\tau}\right)$, sample λ_B^2 from

$$\lambda_B^2 | \{\beta_j, \check{\lambda}_j^2, \check{\tau}_j^2\}_{j=1}^d, a^\tau, c^\tau, e_2 \sim \mathcal{G}\left(\frac{d}{2} + a^\tau, \frac{a^\tau}{4c^\tau} \sum_{j=1}^d \frac{\check{\lambda}_j^2}{\check{\tau}_j^2} \beta_j^2 + e_2\right),$$

and update $\phi^\tau = 2c^\tau / (a^\tau \lambda_B^2)$.

The MCMC scheme in Algorithm 1 is not a full conditional scheme, as several steps are based on partially marginalized distributions. That means that the sampling order matters. For instance, in Step (b), we marginalize w.r.t. $\check{\xi}_1^2, \ldots, \check{\xi}_d^2$, hence we need to update $\check{\xi}_1^2, \ldots, \check{\xi}_d^2$ after sampling a^ξ, before we update c^ξ in Step (d) conditional on $\check{\xi}_1^2, \ldots, \check{\xi}_d^2$. Similarly, due to marginalization in Step (d), we need to update $\check{\kappa}_1^2, \ldots, \check{\kappa}_d^2$, before we update d_2 in Step (f). Furthermore, both Step (b) and Step (d) are based on the marginal prior of κ_B^2, given in (17). Hence, in Step (f), d_2 has to be updated from $d_2 | a^\xi, c^\xi, \kappa_B^2$, before κ_B^2 is updated conditional on d_2.

For a symmetric triple gamma prior, where $a^\xi = c^\xi$, the MCMC scheme in Algorithm 1 has to be modified only slightly. Either $q_a(a^\xi)$ in Step (b) is adjusted and Step (d) is skipped, setting $c^\xi = a^\xi$, or $q_c(c^\xi)$ in Step (d) is adjusted and Step (b) is skipped, setting $a^\xi = c^\xi$. In Appendix B, we provide details in (A11) for the first case and in (A12) for the second case. Similar modifications are needed, if $a^\tau = c^\tau$. All other steps in Algorithm 1 remain the same for $a^\xi = c^\xi$ and/or $a^\tau = c^\tau$.

5. Applications to TVP-VAR-SV Models

5.1. Model

In this section, we consider a generalization of the TVP model (1), where $\boldsymbol{y}_t$ is a m-dimensional time series, observed for $t = 1, \ldots, T$. The time series $\boldsymbol{y}_t$ is assumed to follow a time-varying parameter vector autoregressive model with stochastic volatility (TVP-VAR-SV) of order p:

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{\Phi}_{1,t}\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_{2,t}\boldsymbol{y}_{t-2} + \ldots \boldsymbol{\Phi}_{p,t}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{\Sigma}_t), \tag{28}$$

where c_t is the m-dimensional time-varying intercept, $\mathbf{\Phi}_{j,t}$, for $j = 1, \ldots, p$ is an $m \times m$ matrix of time-varying coefficients, and $\mathbf{\Sigma}_t$ is the time-varying variance covariance matrix of the error term. The TVP-VAR-SV model can be written in a more compact notation as the following TVP model:

$$y_t = (\mathbb{I}_m \otimes \mathbf{x}_t)\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_m\left(\mathbf{0}, \mathbf{\Sigma}_t\right), \tag{29}$$

where $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p}, 1)$ is a row vector of length $mp + 1$ and the time-varying parameter $\boldsymbol{\beta}_t$ is defined as $\boldsymbol{\beta}_t = (\boldsymbol{\beta}_t^{1'}, \ldots, \boldsymbol{\beta}_t^{m'})'$, where $\boldsymbol{\beta}_t^i = (\mathbf{\Phi}_{1,t_{i\bullet}}, \ldots, \mathbf{\Phi}_{p,t_{i\bullet}}, c_{t,i})'$. Here, $\mathbf{\Phi}_{j,t_{i\bullet}}$ denotes the i-th row of the matrix $\mathbf{\Phi}_{j,t}$ and $c_{t,i}$ denotes the i-th element of c_t. Since the influential paper of Primiceri (2005) (see Del Negro and Primiceri (2015) for a corrigendum), this model has become a benchmark for analyzing relationships between macroeconomic variables that evolve over time, see Nakajima (2011), Koop and Korobilis (2013), Eisenstat et al. (2014), Chan and Eisenstat (2016), Feldkircher et al. (2017) and Carriero et al. (2019), among many others.

Following Frühwirth-Schnatter and Tüchler (2008), we use a Cholesky decomposition of the time-varying covariance matrix $\mathbf{\Sigma}_t$, that is $\mathbf{\Sigma}_t = A_t D_t A_t'$, where D_t is a diagonal matrix and A_t is lower unitriangular matrix, see Carriero et al. (2019) and Bitto and Frühwirth-Schnatter (2019) for related models. We denote with $a_{ij,t}$ the element at the i-th row and j-th column of A_t, and with $\sigma_{i,t}^2$ the i-th diagonal element of $D_t = \text{Diag}\left(\sigma_{1,t}^2 \cdots \sigma_{m,t}^2\right)$. In total, we have $m(m-1)/2 + m(mp+1)$ (potentially) time-varying parameters. Using the Cholesky decomposition, we can rewrite the system as:

$$y_t = (\mathbb{I}_m \otimes \mathbf{x}_t)\boldsymbol{\beta}_t + A_t\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}_m\left(\mathbf{0}, D_t\right), \tag{30}$$

where $\boldsymbol{\eta}_t = (\eta_{1,t}, \ldots, \eta_{m,t})^\top$. The idiosyncratic shocks $\eta_{i,t} \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right)$ follow independent SV processes as in (2), with row specific parameters. Specifically, with $h_{i,t} = \log \sigma_{i,t}^2$, we have that the logarithm of the elements of the diagonal matrix D_t follow independent AR(1) processes:

$$h_{i,t} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + v_{i,t}, \quad v_{i,t} \sim \mathcal{N}\left(0, \sigma_{\eta,i}^2\right),$$

for $i = 1, \ldots, m$. Here, μ_i is the mean, ϕ_i is the persistence parameter, and $\sigma_{\eta,i}^2$ is the variance of the ith log-volatility $h_{i,t}$.

It is possible to write the TVP-VAR-SV model (30) as a system of m univariate TVP models as in (1):

$$y_{1,t} = \mathbf{x}_t \boldsymbol{\beta}_t^1 + \eta_{1,t}, \quad \eta_{1,t} \sim \mathcal{N}\left(0, \sigma_{1,t}^2\right),$$
$$y_{2,t} = \mathbf{x}_t \boldsymbol{\beta}_t^2 + a_{21,t}\eta_{1,t} + \eta_{2,t}, \quad \eta_{2,t} \sim \mathcal{N}\left(0, \sigma_{2,t}^2\right),$$
$$y_{3,t} = \mathbf{x}_t \boldsymbol{\beta}_t^3 + a_{31,t}\eta_{1,t} + a_{32,t}\eta_{2,t} + \eta_{3,t}, \quad \eta_{3,t} \sim \mathcal{N}\left(0, \sigma_{3,t}^2\right),$$
$$\cdots$$
$$y_{m,t} = \mathbf{x}_t \boldsymbol{\beta}_t^m + a_{m1,t}\eta_{1,t} + \ldots + a_{m,m-1,t}\eta_{m-1,t} + \eta_{m,t}, \quad \eta_{m,t} \sim \mathcal{N}\left(0, \sigma_{m,t}^2\right).$$

Note that for $i > 1$, the i-th equation of this system is a TVP model where the residuals of the preceding $i - 1$ equations are added as explanatory variables:

$$y_{i,t} = \mathbf{x}_t \boldsymbol{\beta}_t^i + \sum_{j=1}^{i-1} a_{ij,t}\eta_{j,t} + \eta_{i,t}, \quad \eta_{i,t} \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right),$$

and all time-varying parameters follow a random walk as in the TVP model (1):

$$\beta_{j,t}^i = \beta_{j,t-1}^i + v_{ij,t}, \quad v_{ij,t} \sim \mathcal{N}\left(0, \theta_{ij}^\beta\right), \quad \text{for} \quad i = 1,\ldots,m, \quad \text{and} \quad j = 1,\ldots,mp+1,$$

$$a_{ij,t} = a_{ij,t-1} + w_{ij,t}, \quad w_{ij,t} \sim \mathcal{N}\left(0, \theta_{ij}^a\right), \quad \text{for} \quad i = 1,\ldots,m, \quad \text{and} \quad j = 1,\ldots,i-1,$$

with initial values $\beta_{j,0}^i \sim \mathcal{N}\left(\beta_{ij}^\beta, \theta_{ij}^\beta\right)$ and $a_{ij,0} \sim \mathcal{N}\left(\beta_{ij}^a, \theta_{ij}^a\right)$. Here, $\beta_{j,t}^i$ denotes the jth element of the vector $\boldsymbol{\beta}_t^i$.

To achieve shrinkage for each VAR coefficient $\beta_{j,t}^i$ as well as for each Cholesky factor $a_{ij,t}$, we proceed as in Section 2 and introduce shrinkage priors for the initial expectations β_{ij}^β and β_{ij}^a as well as the variances θ_{ij}^β and θ_{ij}^a. We do this independently for each equation of the system. Within each equation, the β_{ij}^βs and β_{ij}^as are assumed to follow independent shrinkage priors to allow for flexibility in the prior structure, and similarly for θ_{ij}^β and θ_{ij}^a:

$$\beta_{ij}^x \sim \mathcal{N}\left(0, \phi_i^{\tau,x}\check{\tau}_{ij}^{x,2}/\check{\lambda}_{ij}^{x,2}\right), \quad \check{\tau}_{ij}^{x,2} \sim \mathcal{G}\left(a_i^{\tau,x}, 1\right), \quad \check{\lambda}_{ij}^{x,2} \sim \mathcal{G}\left(c_i^{\tau,x}, 1\right), \quad \phi_i^{\tau,x} = 2c_i^{\tau,x}/(\lambda_{B,i}^{x,2}a_i^{\tau,x}), \quad (31)$$

$$\sqrt{\theta_{ij}^x} \sim \mathcal{N}\left(0, \phi_i^{\xi,x}\check{\xi}_{ij}^{x,2}/\check{\kappa}_{ij}^{x,2}\right), \quad \check{\xi}_{ij}^{x,2} \sim \mathcal{G}\left(a_i^{\xi,x}, 1\right), \quad \check{\kappa}_{ij}^{x,2} \sim \mathcal{G}\left(c_i^{\xi,x}, 1\right), \quad \phi_i^{\xi,x} = 2c_i^{\xi,x}/(\kappa_{B,i}^{x,2}a_i^{\xi,x}),$$

where $x = \beta$ for the VAR-coefficients and $x = a$ for the elements of A_t. Following Section 2.4, the priors for the global shrinkage parameters in the ith equation read

$$\lambda_{B,i}^{x,2}|a_i^{\tau,x}, c_i^{\tau,x} \sim \mathrm{F}\left(2a_i^{\tau,x}, 2c_i^{\tau,x}\right), \quad 2a_i^{\tau,x} \sim \mathcal{B}\left(\alpha_{a^\tau}, \beta_{a^\tau}\right), \quad 2c_i^{\tau,x} \sim \mathcal{B}\left(\alpha_{c^\tau}, \beta_{c^\tau}\right), \quad (32)$$

$$\kappa_{B,i}^{x,2}|a_i^{\xi,x}, 2c_i^{\xi,x} \sim \mathrm{F}\left(2a_i^{\xi,x}, 2c_i^{\xi,x}\right), \quad 2a_i^{\xi,x} \sim \mathcal{B}\left(\alpha_{a^\xi}, \beta_{a^\xi}\right), \quad 2c_i^{\xi,x} \sim \mathcal{B}\left(\alpha_{c^\xi}, \beta_{c^\xi}\right).$$

5.2. A Brief Sketch of the TVP-VAR-SV MCMC Algorithm

Our algorithm exploits the aforementioned unitriangular decomposition to estimate the model parameters equation-by-equation. Due to the prior structure introduced in (31), the estimation of $\boldsymbol{\beta}_t^i$ and the $a_{ij,t}$'s is separated into two blocks, with the algorithm cycling through the m equations, alternating between sampling $\boldsymbol{\beta}_t^i$ conditional on Σ_t and sampling the $a_{ij,t}$s and $d_{i,t}$s conditional on the VAR coefficients $\boldsymbol{\beta}_t^i$. Given a set of initial values, the algorithm repeats the following steps:

Algorithm 2. MCMC inference for TVP-VAR-SV models under the triple gamma prior.

Choose starting values for all global and local shrinkage parameters in prior (31) for each equation and repeat the following steps:

For $i = 1,\ldots,m$, update all the unknowns in the ith equation:

(a) *Conditional on A_t and D_t, create $\check{y}_{i,t} = y_{i,t} - \sum_{j=1}^{i-1} a_{ij,t}\eta_{j,t}$ and define the following TVP model:*

$$\check{y}_{i,t} = \mathbf{x}_t\boldsymbol{\beta}_t^i + \eta_{i,t}, \quad \eta_{i,t} \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right).$$

Apply Algorithm 1 (sans the step for the variance of the observation equation) to this univariate TVP model, to draw from the conditional posterior distribution of the time-varying VAR-coefficients $\boldsymbol{\beta}_t^i$, for $t = 0,\ldots,T$, their initial expectations β_{ij}^β, the process variances θ_{ij}^β, the local shrinkage parameters $\check{\tau}_{ij}^{\beta,2}, \check{\lambda}_{ij}^{\beta,2}, \check{\xi}_{ij}^{\beta,2}, \check{\kappa}_{ij}^{\beta,2}$, as well as the global shrinkage parameters $\lambda_{B,i}^{\beta,2}, \kappa_{B,i}^{\beta,2}, a_i^{\tau,\beta}, c_i^{\tau,\beta}, a_i^{\xi,\beta}$, and $c_i^{\xi,\beta}$.

(b) *For $i > 1$, create $y_{i,t}^\star = y_{i,t} - \mathbf{x}_t\boldsymbol{\beta}_t^i$, conditional on $\boldsymbol{\beta}_t^i$, and define the following TVP model:*

$$y_{i,t}^\star = \sum_{j=1}^{i-1} a_{ij,t}\eta_{j,t} + \eta_{i,t}, \quad \eta_{i,t} \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right),$$

where the residuals from the previous $i − 1$ equations, $(\eta_{1,t}, \ldots, \eta_{i−1,t})$, are used as explanatory variables and no intercept is present. Apply Algorithm 1 to this univariate TVP model, to sample the volatilities $\sigma^2_{i,t}$ and the time-varying coefficients $a_{ij,t}$ in the ith row of A_t for $t = 0, \ldots, T$ from the respective conditional posteriors, as well as the initial expectations β^a_{ij}, the process variances θ^a_{ij}, the local shrinkage parameters $\check{\tau}^{a,2}_{ij}, \check{\lambda}^{a,2}_{ij}, \check{\xi}^{a,2}_{ij}, \check{\kappa}^{a,2}_{ij}$ and the global shrinkage parameters $\lambda^{a,2}_{B,i}, \kappa^{a,2}_{B,i}, a^{\tau,a}_i, c^{\tau,a}_i, a^{\xi,a}_i,$ and $c^{\xi,a}_i$.

In the following applications, we run our algorithm for $M = 200,000$ iterations, discarding the first $100,000$ iterations as burn-in, and then keeping the output of one every 100 iterations.

5.3. Illustrative Example with Simulated Data

To illustrate the merit of our methodology in the context of TVP-VAR-SV models, we simulate data from two TVP-VAR-SV models with $T = 200$ points in time, $p = 1$ lags and $m = 7$ equations, with varying degrees of sparsity. In the dense regime, approximately 30% of the values of β and θ (here referring to the means of the initial states and the variances of the innovations as defined in Section 2, respectively) are truly zero, while in the sparse regime approximately 90% are truly zero. We show results for the triple gamma prior, the Horseshoe prior, the double gamma and the Lasso.

Regarding the priors on the hyperparameters, we use prior (32) with $\alpha_{a^\tau} = \alpha_{c^\tau} = \alpha_{a^\xi} = \alpha_{c^\xi} = 1$ and $\beta_{a^\tau} = \beta_{c^\tau} = \beta_{a^\xi} = \beta_{c^\xi} = 6$ for the triple gamma. The probability density function of the corresponding beta prior is monotonically increasing, with a maximum at 0.5. This prior places positive mass in a neighborhood of the Horseshoe, but allows for more flexibility. In practice, placing a prior on the spike and slab parameters of the triple gamma, instead of fixing them to 0.5 as in the Horseshoe, allows us to learn the shrinkage profile from the data, including asymmetric profiles.

We assume that the global shrinkage parameters $\lambda^{\beta,2}_{B,i}, \kappa^{\beta,2}_{B,i}, \lambda^{a,2}_{B,i},$ and $\kappa^{a,2}_{B,i}$ follow a $F(1,1)$ distribution for the Horseshoe prior which corresponds to the prior in Carvalho et al. (2009) and a $\mathcal{G}(0.001, 0.001)$ distribution for the Lasso and the double gamma prior, as suggested in Belmonte et al. (2014) and Bitto and Frühwirth-Schnatter (2019). Concerning the spike parameters $a^{\tau,a}_i, a^{\xi,a}_i, a^{\tau,\beta}_i,$ and $a^{\xi,\beta}_i$ of the double gamma, we employ a rescaled beta prior to force them to be smaller than 0.5. Specifically, we use a $B(4,6)$ prior which places most of its mass between 0.05 and 0.4, a range that Bitto and Frühwirth-Schnatter (2019) have found to induce desirable shrinkage characteristics.
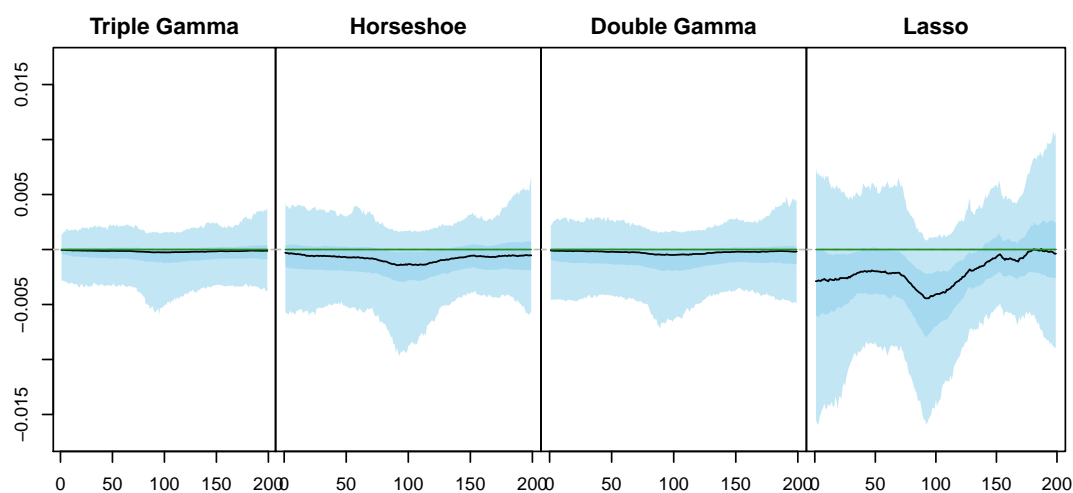

Figure 5. Posterior path against time for a constant non-significant parameter $\beta^i_{j,t}$ in the sparse regime.

Figure 5 shows the posterior path of a permanently non-significant state, that is a state where the true $\beta^i_{j,t} = 0$ for $t = 1, \ldots, T$, in the sparse regime. The entire set of states for the triple gamma prior

can be found in Appendix C. Note that, while the zero line is contained in the 95% posterior credible interval for all priors, said interval is thinner under the triple gamma prior and the double gamma prior than under the Lasso and the Horseshoe prior.

We calculate the posterior inclusion probabilities based on the thresholding approach introduced in Section 3.2, comparing the triple gamma prior to widely used special cases. In a variance selection context, the posterior inclusion probabilities reflect the uncertainty on whether a state should be time varying or constant over time. Figure 6 shows the posterior inclusion probabilities for the variance of the innovations (θ_{ij}^{β}'s) under four different shrinkage priors, for the sparse and the dense scenario, respectively. The cells are shaded in gray when the corresponding true state parameter is time-varying ($\theta_{ij}^{\beta} \neq 0$), while the background is white when the corresponding true state parameter is not time-varying ($\theta_{ij}^{\beta} = 0$). In this simulated example, the posterior inclusion probabilities under the triple gamma prior are consistently higher for the variances that are actually different from 0, even when they are very small. This outcome is in line with the analytical results derived in Section 2.2, which show that the tails of the triple gamma prior are heavier than those of the other priors.

Figure 6. Posterior inclusion probability for the θ_{ij}^{β}'s in the sparse and dense regime, under the triple gamma prior, the Horseshoe prior, the Lasso prior and the double gamma prior. The true values of the θ_{ij}^{β}'s are reported in each cell.

5.4. Modeling Area Macroeconomic and Financial Variables in the Euro Area

Our application investigates a subset of the area wide model of the European Union of Fagan et al. (2005), which comprises quarterly macroeconomic data spanning from 1970 to 2017. We include seven of the variables present in the dataset, namely real output (YER), prices (YED), short-term interest rate (STN), investment (ITR), consumption (PCR), exchange rate (EEN) and unemployment (URX). A more detailed description of the data and the transformations performed to make the time series stationary can be found in Table A1 in Appendix D. To stay in line with the literature, for example, Feldkircher et al. (2017), we estimate a TVP-VAR-SV model with $p = 2$ lags on all endogenous variables. The hyperparameter choices are the same as in Section 5.3. As in the example with simulated data, we run the algorithm for $M = 200,000$ iterations, discarding the first $100,000$ iterations as burn-in, and then keeping the output of one every 100 iterations.

Figures 7 and 8 display the posterior inclusion probabilities for the means of the initial states and the innovation variances of the VAR coefficients, respectively. A few things about Figure 7 are noteworthy. First, the posterior inclusion probabilities on the diagonal, meaning those belonging to the parameter of each equation's own autoregressive term, appear to be those that are the highest, while off diagonal elements are more likely to be excluded. Second, the equation for the short-term

interest rate is characterized by a large amount of parameters with a high inclusion probability, across all priors. Third, the first lag tends to have higher posterior inclusion probabilities than the second lag, which is in line with the literature. In most cases, the triple gamma prior can be seen to have either the largest or the smallest posterior inclusion probability compared to the other priors. This can be seen as a reflection of the fact that the triple gamma prior places more mass on the edges of the shrinkage profile, as illustrated in Section 3.

Now, we shift our focus to the posterior inclusion probabilities for the θ_{ij}^{β}'s plotted in Figure 8. Compared to the means of the inital states, almost all inclusion probabilities are essentially zero. This lack of variability is unsurprising, as it is well known (see, e.g., Feldkircher et al. (2017)) that stochastic volatility in a TVP-VAR model for macroeconomic variables can explain a large part of the variability in the data. However, the triple gamma prior appears to allow posterior distributions that place slightly more mass on models with some time variation, in particular with respect to the financial variables.

Figures 9 and 10 display the posterior median of β_{ij}^{β} and $\left| \sqrt{\theta}_{ij}^{\beta} \right|$, respectively. Here the triple gamma can be seen to be quite conservative, both in terms of which parameters to include, as well as their magnitude. In particular the medians of the $\left| \sqrt{\theta}_{ij}^{\beta} \right|$ are interesting, as they are closest to zero under the triple gamma prior, despite having the highest posterior inclusion probabilities among all considered priors.

In Figures A3 and A4 in Appendix D, all the posterior paths of $\Phi_{1,t}$ and $\Phi_{2,t}$ under the triple gamma prior are shown.
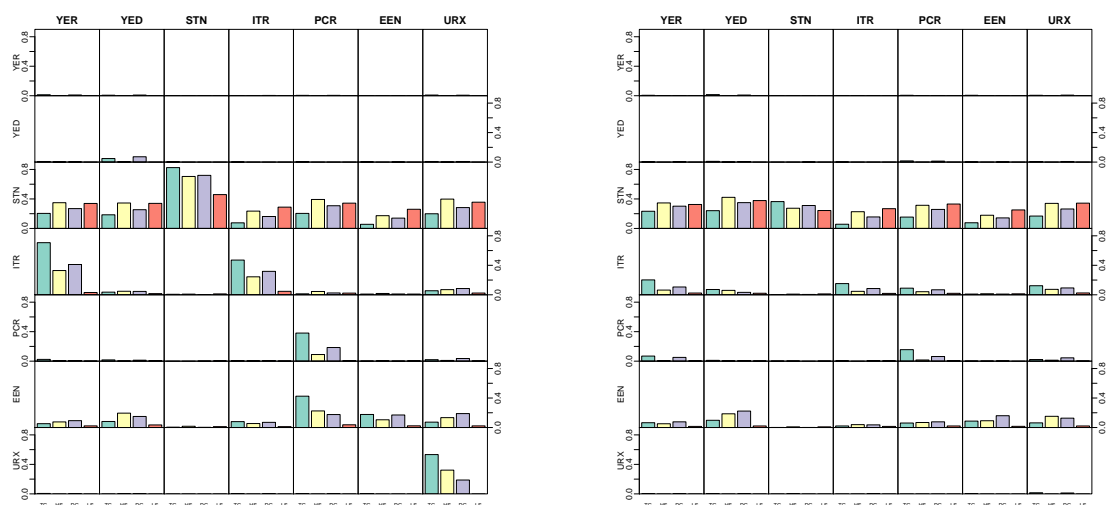

Figure 7. Posterior inclusion probability for state parameters β_{ij}^{β} associated with the first lag (on the **left**) and with the second lag (on the **right**), for the Euro Area data under the triple gamma prior, the Horseshoe prior, the double gamma prior and the Lasso prior.
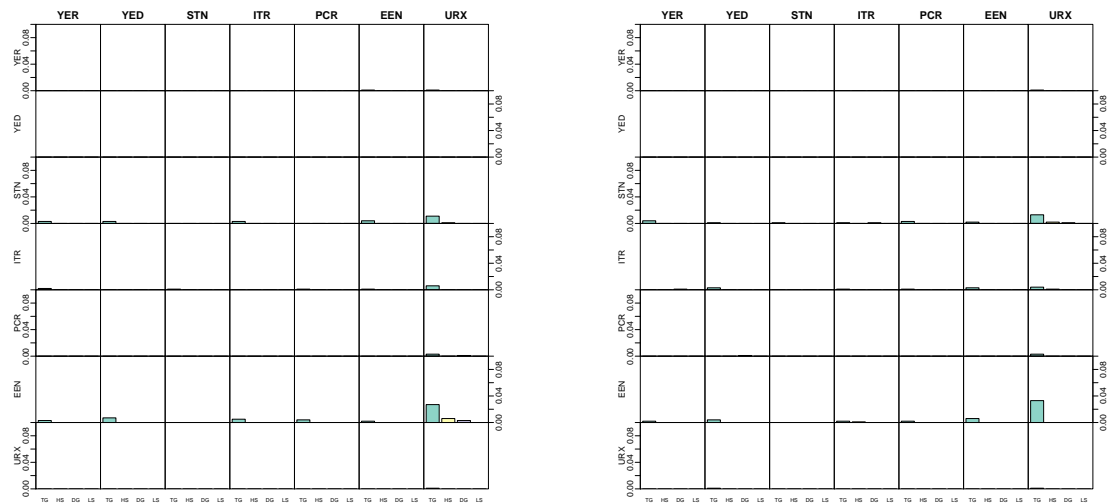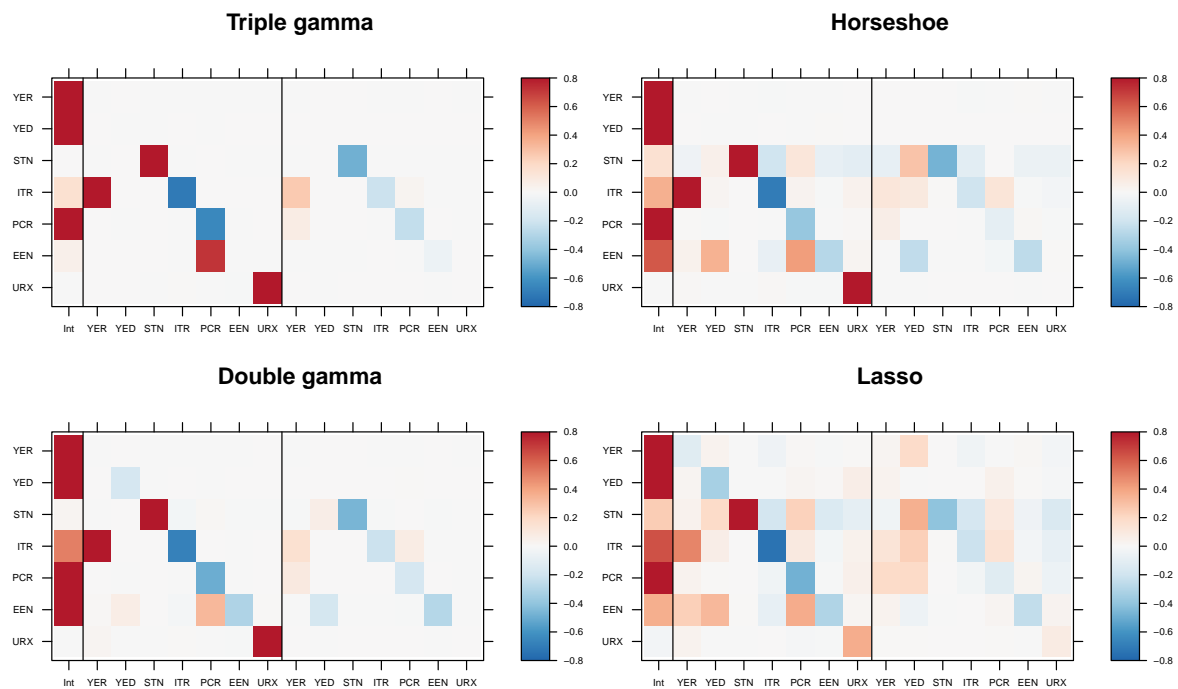
Figure 8. Posterior inclusion probability for θ_{ij}^{β}'s associated with the first lag on the **left** and with the second lag on the **right**, for the Euro Area data under the triple gamma prior, the Horseshoe prior, the double gamma prior and the Lasso prior.
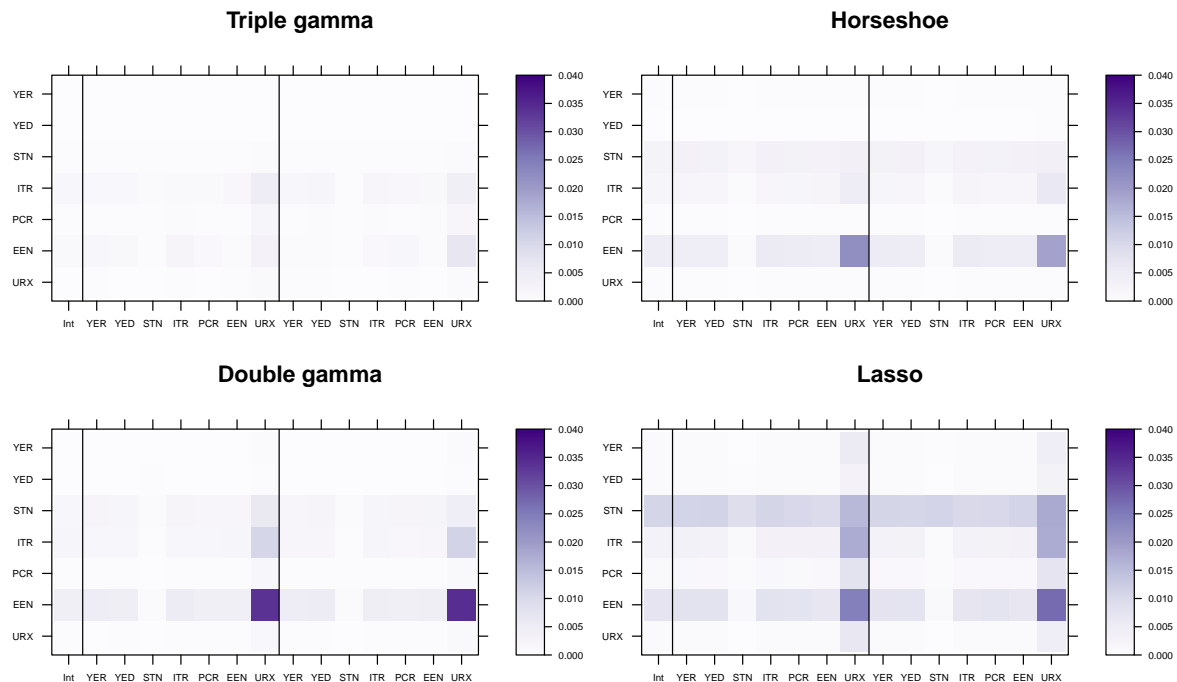

Figure 9. Posterior median of β_{ij}^{β} under the triple gamma, Horseshoe, double gamma and Lasso for the Euro area model. The vertical lines delimit the intercept, first and second lag, respectively.

Figure 10. Posterior median of $\left|\sqrt{\bar{\theta}_{ij}^{\beta}}\right|$ under the triple gamma, Horseshoe, double gamma and Lasso for the Euro area model. The vertical lines delimit the intercept, first and second lag, respectively.

6. Conclusions

In the present paper, shrinkage for time-varying parameter (TVP) models was investigated within a Bayesian framework with the goal to automatically reduce time-varying parameters to static ones, if the model is overfitting. This goal was achieved by suggesting the triple gamma prior as a new shrinkage priors for the process variances of varying coefficients, extending previous work using spike-and-slab priors, the Bayesian Lasso, or the double gamma prior. The triple gamma prior is related to the normal-gamma-gamma prior applied for variable selection in highly structured regression models (Griffin and Brown 2017). It contains the well-known Horseshoe prior as a special case, however it is more flexible, with two shape parameters that control concentration at zero and the tail behaviour. This leads to a BMA-type behaviour which allows not only variance shrinkage, but also variance selection.

In our application, we considered time-varying parameter VAR models with stochastic volatility. Overall, our findings suggest that the family of triple gamma priors introduced in this paper for sparse TVP models is successful in avoiding overfitting, if coefficients are, indeed, static or even insignificant. The framework developed in this paper is very general and holds the promise to be useful for introducing sparsity in other TVP and state space models in many different settings.

A number of extensions seem to be worth pursuing. First of all, the triple gamma prior is relevant not only for TVP models, but for any model containing variance parameters such as random-effect models or Bayesian p-splines models (Scheipl and Kneib 2009). Second, in particular, in ultra-sparse settings, modifications of the triple gamma prior seem sensible. Currently, the hyperprior for the global shrinkage parameter of the triple gamma prior is selected in a way that it implies a uniform prior on "model size". A generalization of Theorem 3 would allow the choice of hyper priors that induce higher sparsity. Furthermore, in the variable selection literature, special priors such as the Horseshoe+ (Bhadra et al. 2017a) were suggested for very sparse, ultra-sparse high dimensional settings. Exploiting

once more the non-centered parametrization of a state space model, it is straightforward to extend this prior to variance selection using following hierarchical representation:

$$\sqrt{\theta_j}|\kappa_j^2, \xi_j^2 \sim \mathcal{N}\left(0, \frac{2}{\kappa_B^2}\kappa_j^2\xi_j^2\right), \quad \kappa_j \sim t_1, \quad \xi_j \sim t_1.$$

We leave these extensions for future research.

Finally, an important limitation of our approach is that shrinking a variance toward zero implies that a coefficient is fixed over the entire observation period of the time series. In future research we will investigate dynamic shrinkage priors (Kalli and Griffin 2014; Kowal et al. 2019; Ročková and McAlinn 2020) where coefficients can be both fixed and dynamic.

Author Contributions: The authors contributed equally to the work. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that there is no conflict of interest.

Appendix A. Proofs

Proof of Theorem 1. To proof Part (a), rewrite prior (6) in the following way by rescaling ξ_j^2 and κ_j^2:

$$\sqrt{\theta_j}|\tilde{\xi}_j^2, \tilde{\kappa}_j^2, \kappa_B^2 \sim \mathcal{N}\left(0, \frac{2}{\kappa_B^2}\frac{\tilde{\xi}_j^2}{\tilde{\kappa}_j^2}\right), \quad \tilde{\xi}_j^2|a^\xi \sim \mathcal{G}\left(a^\xi, a^\xi\right), \quad \tilde{\kappa}_j^2|c^\xi \sim \mathcal{G}\left(c^\xi, c^\xi\right), \tag{A1}$$

and use the fact that in (A1) the random variable $\psi_j^2 = \tilde{\xi}_j^2/\tilde{\kappa}_j^2$ follows the F-distribution:

$$\psi_j^2 = \frac{\tilde{\xi}_j^2}{\tilde{\kappa}_j^2} \sim \frac{\mathcal{G}\left(a^\xi, a^\xi\right)}{\mathcal{G}\left(c^\xi, c^\xi\right)} =_d \mathrm{F}\left(2a^\xi, 2c^\xi\right),$$

where $p(\psi_j^2)$ is given by:

$$p(\psi_j^2) = \frac{1}{B(a^\xi, c^\xi)}\left(\frac{a^\xi}{c^\xi}\psi_j^2\right)^{a^\xi-1}\left(1 + \frac{a^\xi}{c^\xi}\psi_j^2\right)^{-(a^\xi+c^\xi)}. \tag{A2}$$

This yields (8).

Using that $\eta_j = 1/\psi_j^2 \sim \mathrm{F}\left(2c^\xi, 2a^\xi\right)$, we obtain from (8) that

$$p(\sqrt{\theta_j}|\kappa_B^2, a^\xi, c^\xi) = \frac{\sqrt{\kappa_B^2}(c^\xi)^{c^\xi}}{\sqrt{4\pi}(a^\xi)^{c^\xi}B(a^\xi, c^\xi)}\int_0^\infty \exp\left(-\frac{\theta_j\kappa_B^2\eta_j}{4}\right)\eta_j^{c^\xi-\frac{1}{2}}\left(1 + \frac{c^\xi\eta_j}{a^\xi}\right)^{-(a^\xi+c^\xi)}\,d\eta_j.$$

A change of variable with $y_j = c^\xi\eta_j/a^\xi$ proves Part (b):

$$p(\sqrt{\theta_j}|\phi^\xi, a^\xi, c^\xi) = \frac{1}{\sqrt{2\pi\phi^\xi}B(a^\xi, c^\xi)}\int_0^\infty \exp\left(-\frac{\theta_j}{2\phi^\xi}y_j\right)y_j^{c^\xi-\frac{1}{2}}\left(1 + y_j\right)^{-(a^\xi+c^\xi)}\,dy_j$$

$$= \frac{\Gamma(c^\xi + \frac{1}{2})}{\sqrt{2\pi\phi^\xi}B(a^\xi, c^\xi)}U\left(c^\xi + \frac{1}{2}, \frac{3}{2} - a^\xi, \frac{\theta_j}{2\phi^\xi}\right),$$

where $\phi^\xi = \frac{2c^\xi}{\kappa_B^2 a^\xi}$. □

Proof of Theorem 2. Using Abramowitz and Stegun (1973, 13.5.8), we obtain for a and $1 < b < 2$ fixed that $U(a, b, z)$ behaves for small z as:

$$U(a, b, z) = \frac{\Gamma(b-1)}{\Gamma(a)} z^{1-b} + O(1).$$

Since $b = 3/2 - a^\xi$ in the expression for $p(\sqrt{\theta_j}|\phi^\xi, a^\xi, c^\xi)$ given in (9), the condition $1 < b < 2$ is equivalent to $0 < a^\xi < 0.5$ and this proves Part (a):

$$p(\sqrt{\theta_j}|\phi^\xi, a^\xi, c^\xi) = \frac{\Gamma(\frac{1}{2} - a^\xi)}{\sqrt{\pi}(2\phi^\xi)^{a^\xi} B(a^\xi, c^\xi)} \left(\frac{1}{\sqrt{\theta_j}}\right)^{1-2a^\xi} + O(1).$$

For $b = 1$ we obtain from Abramowitz and Stegun (1973, 13.5.9) that $U(a, b, z)$ behaves for small z as follows:

$$U(a, b, z) = -\frac{1}{\Gamma(a)}(\log z + \psi(a)) + O(|z \log z|),$$

where $\psi(\cdot)$ is the digamma function. Since $b = 1$ is equivalent with $a^\xi = 0.5$, this proves Part (b):

$$p(\sqrt{\theta_j}|\phi^\xi, a^\xi, c^\xi) = \frac{1}{\sqrt{2\pi\phi^\xi}B(a^\xi, c^\xi)}\left(-\log\theta_j + \log(2\phi^\xi) - \psi(c^\xi + \tfrac{1}{2})\right) + O(|\theta_j \log\theta_j|).$$

Using formulas 13.5.10-13.5.12 in Abramowitz and Stegun (1973), we obtain for a and $b < 1$ fixed that $U(a, b, z)$ behaves for small z as follows:

$$U(a, b, z) = \begin{cases} \dfrac{\Gamma(1-b)}{\Gamma(1+a-b)} + O(z^{1-b}), & 0 < b < 1, \\[2mm] \dfrac{1}{\Gamma(1+a)} + O(|z \log z|), & b = 0, \\[2mm] \dfrac{\Gamma(1-b)}{\Gamma(1+a-b)} + O(|z|), & b < 0. \end{cases}$$

Since $O(z^{1-b})$ with $b < 1$, $O(|z \log z|)$ and $O(|z|)$ converge to 0 as $z \to 0$, we obtain:

$$\lim_{z \to 0} U(a, b, z) = \frac{\Gamma(1-b)}{\Gamma(1+a-b)}.$$

This proves Part (c) as condition $b < 1$ is equivalent to $a^\xi > 0.5$:

$$\lim_{\sqrt{\theta_j} \to 0} p(\sqrt{\theta_j}|\phi^\xi, a^\xi, c^\xi) = \frac{\Gamma(c^\xi + \frac{1}{2})}{\sqrt{2\pi\phi^\xi}B(a^\xi, c^\xi)} \lim_{z \to 0} U\left(c^\xi + \frac{1}{2}, \frac{3}{2} - a^\xi, z\right) = \frac{\Gamma(c^\xi + \frac{1}{2})\Gamma(a^\xi - \frac{1}{2})}{\sqrt{2\pi\phi^\xi}B(a^\xi, c^\xi)\Gamma(a^\xi + c^\xi)}.$$

Finally, using Abramowitz and Stegun (1973, 13.1.8), we obtain as $z \to \infty$:

$$U(a, b, z) = z^{-a}\left[1 + O\left(\frac{1}{z}\right)\right].$$

Therefore as $\sqrt{\theta_j} \to \infty$

$$p(\sqrt{\theta_j}|\phi^\xi, a^\xi, c^\xi) = \frac{\Gamma(c^\xi + \frac{1}{2})(2\phi^\xi)^{c^\xi}}{\sqrt{\pi}B(a^\xi, c^\xi)}\left(\frac{1}{\sqrt{\theta_j}}\right)^{2c^\xi + 1}\left[1 + O\left(\frac{1}{\theta_j}\right)\right].$$

$\square$

Proof of Lemma 1. To derive representation (a), integrate (A1) with respect to $\check{\kappa}_j^2$, using the common normal-scale mixture representation of the Student-t distribution. Representation (b) is obtained from (10) by rescaling. Representation (c) is obtained from (A1) by rescaling $\tilde{\zeta}_j^2$ and $\tilde{\kappa}_j^2$. Finally, by defining $\tilde{\psi}_j^2 = \frac{a^{\tilde{\xi}}}{c^{\tilde{\xi}}} \psi_j^2$, representation (d) follows immediately from (8) and (A2). □

Proof of Theorem 3. The equivalence of (23) and (24) follows immediately from

$$\phi^{\tilde{\xi}} = (c^{\tilde{\xi}}/a^{\tilde{\xi}})(2/\kappa_B^2) \sim \mathcal{BP}\left(c^{\tilde{\xi}}, a^{\tilde{\xi}}\right),$$

since $2/\kappa_B^2 \sim \mathrm{F}\left(2c^{\tilde{\xi}}, 2a^{\tilde{\xi}}\right)$. In addition, (24) implies that

$$\frac{\phi^{\tilde{\xi}}}{1+\phi^{\tilde{\xi}}} \sim \mathcal{B}\left(c^{\tilde{\xi}}, a^{\tilde{\xi}}\right).$$

Using representations (13) and (14) of the tripe gamma prior, we can show:

$$\rho_j < 0.5 \quad \Leftrightarrow \quad \xi_j^2 = \frac{1}{\rho_j} - 1 > 1 \quad \Leftrightarrow \quad \phi^{\tilde{\xi}}\tilde{\psi}_j^2 > 1 \quad \Leftrightarrow \quad \frac{1}{1+\tilde{\psi}_j^2} < \frac{\phi^{\tilde{\xi}}}{1+\phi^{\tilde{\xi}}},$$

where $\tilde{\psi}_j^2 \sim \mathcal{BP}\left(a^{\tilde{\xi}}, c^{\tilde{\xi}}\right)$ and, consequently,

$$\frac{\tilde{\psi}_j^2}{1+\tilde{\psi}_j^2} \sim \mathcal{B}\left(a^{\tilde{\xi}}, c^{\tilde{\xi}}\right) \quad \Leftrightarrow \quad \frac{1}{1+\tilde{\psi}_j^2} \sim \mathcal{B}\left(c^{\tilde{\xi}}, a^{\tilde{\xi}}\right).$$

Hence, $\pi^{\tilde{\xi}} = \Pr(\rho_j < 0.5) = F_X(Y)$, where F_X is the cdf of a random variable $X \sim \mathcal{B}\left(c^{\tilde{\xi}}, a^{\tilde{\xi}}\right)$ and the random variable $Y \sim \mathcal{B}\left(c^{\tilde{\xi}}, a^{\tilde{\xi}}\right)$ arises from the same distribution. It follows immediately that $\pi^{\tilde{\xi}} \sim \mathcal{U}[0, 1]$. □

Appendix B. Details on the MCMC scheme

In Step (b),

$$p(a^{\tilde{\xi}}|\mathbf{z}_{-a^{\tilde{\xi}}}, \mathbf{y}) \propto \prod_{j=1}^{d} p(\sqrt{\theta_j}|\check{\kappa}_j^2, \phi^{\tilde{\xi}})\, p(\kappa_B^2|a^{\tilde{\xi}}, c^{\tilde{\xi}})\, p(a^{\tilde{\xi}}),$$

where $p(\kappa_B^2|a^{\tilde{\xi}}, c^{\tilde{\xi}})$ is given by:

$$p(\kappa_B^2|a^{\tilde{\xi}}, c^{\tilde{\xi}}) = \frac{1}{2^{a^{\tilde{\xi}}} B(a^{\tilde{\xi}}, c^{\tilde{\xi}})} \left(\frac{a^{\tilde{\xi}}}{c^{\tilde{\xi}}}\kappa_B^2\right)^{a^{\tilde{\xi}}-1} \left(1 + \frac{a^{\tilde{\xi}}}{2c^{\tilde{\xi}}}\kappa_B^2\right)^{-(a^{\tilde{\xi}}+c^{\tilde{\xi}})}.$$

Therefore,

$$\begin{aligned}
p(a^{\tilde{\xi}}|\mathbf{z}_{-a^{\tilde{\xi}}}, \mathbf{y}) \propto\ & \frac{2^{-da^{\tilde{\xi}}}}{\Gamma(a^{\tilde{\xi}})^d}(a^{\tilde{\xi}})^{d(a^{\tilde{\xi}}+1/2)/2}\left(\frac{\kappa_B^2}{c^{\tilde{\xi}}}\right)^{da^{\tilde{\xi}}/2} \\
& \cdot \left(\prod_{j=1}^{d}\check{\kappa}_j^2\theta_j\right)^{a^{\tilde{\xi}}/2} \prod_{j=1}^{d} K_{a^{\tilde{\xi}}-1/2}\left(\sqrt{\check{\kappa}_j^2\kappa_B^2 a^{\tilde{\xi}}|\theta_j|/c^{\tilde{\xi}}}\right) \\
& \cdot \frac{1}{2^{a^{\tilde{\xi}}} B(a^{\tilde{\xi}}, c^{\tilde{\xi}})}\left(\frac{a^{\tilde{\xi}}}{c^{\tilde{\xi}}}\kappa_B^2\right)^{a^{\tilde{\xi}}-1}\left(1 + \frac{a^{\tilde{\xi}}}{2c^{\tilde{\xi}}}\kappa_B^2\right)^{-(a^{\tilde{\xi}}+c^{\tilde{\xi}})}(2a^{\tilde{\xi}})^{\alpha_{a^{\tilde{\xi}}}-1}(1-2a^{\tilde{\xi}})^{\beta_{a^{\tilde{\xi}}}-1}.
\end{aligned} \tag{A3}$$

Hence, $\log q_a(a^{\breve{\xi}})$ is given by (using $\Gamma(a^{\breve{\xi}}) = \Gamma(a^{\breve{\xi}}+1)/a^{\breve{\xi}}$):

$$
\begin{aligned}
\log q_a(a^{\breve{\xi}}) \quad = \quad & a^{\breve{\xi}}\left(-d\log 2 + \frac{d}{2}\log\kappa_B^2 - \frac{d}{2}\log c^{\breve{\xi}} + \frac{1}{2}\sum_{j=1}^d \log\breve{\kappa}_j^2 + \frac{1}{2}\sum_{j=1}^d \log\theta_j\right) \\
& + \frac{5}{4}d\log a^{\breve{\xi}} + d\frac{a^{\breve{\xi}}}{2}\log a^{\breve{\xi}} - d\log\Gamma(a^{\breve{\xi}}+1) \\
& + \sum_{j=1}^d \log K_{a^{\breve{\xi}}-1/2}\left(\sqrt{\breve{\kappa}_j^2\kappa_B^2 a^{\breve{\xi}}|\theta_j|/c^{\breve{\xi}}}\right) \qquad \text{(prior on } \theta_j) \\
& - \log B(a^{\breve{\xi}},c^{\breve{\xi}}) + a^{\breve{\xi}}\left(\log a^{\breve{\xi}} + \log(\tfrac{\kappa_B^2}{2c^{\breve{\xi}}})\right) - \log a^{\breve{\xi}} - (a^{\breve{\xi}}+c^{\breve{\xi}})\log\left(1 + \frac{a^{\breve{\xi}}\kappa_B^2}{2c^{\breve{\xi}}}\right) \quad \text{(prior on } \kappa_B^2) \\
& + (\alpha_{a^{\breve{\xi}}}-1)\log(2a^{\breve{\xi}}) - (\beta_{a^{\breve{\xi}}}-1)\log(1-2a^{\breve{\xi}}) \qquad \text{(prior on } a^{\breve{\xi}}) \\
& + \log a^{\breve{\xi}} + \log(0.5 - a^{\breve{\xi}}) \qquad \text{(change of variable)}
\end{aligned}
\tag{A4}
$$

In Step (c),

$$
\begin{aligned}
p(\breve{\xi}_j^2 | z_{-\breve{\xi}_j^2}, y) \quad & \propto \quad p(\sqrt{\theta_j}|\breve{\xi}_j^2,\breve{\kappa}_j^2,\phi^{\breve{\xi}})p(\breve{\xi}_j^2|a^{\breve{\xi}}) \\
& \propto \quad (\breve{\xi}_j^2)^{-1/2}\exp\left\{-\frac{\breve{\kappa}_j^2}{2\phi^{\breve{\xi}}\breve{\xi}_j^2}\theta_j\right\}\cdot(\breve{\xi}_j^2)^{a^{\breve{\xi}}-1}\exp\left\{-\breve{\xi}_j^2\right\} \\
& = \quad (\breve{\xi}_j^2)^{a^{\breve{\xi}}-1/2-1}\exp\left\{-\frac{1}{2}\left(\frac{\breve{\kappa}_j^2\theta_j}{\phi^{\breve{\xi}}}\frac{1}{\breve{\xi}_j^2} + 2\breve{\xi}_j^2\right)\right\},
\end{aligned}
\tag{A5}
$$

which is equal to the GIG-distribution given in (25).[5]
In Step (d),

$$
\begin{aligned}
p(c^{\breve{\xi}}|z_{-c^{\breve{\xi}}}, y) \quad & \propto \quad \prod_{j=1}^d p(\sqrt{\theta_j}|\breve{\xi}_j^2,c^{\breve{\xi}},\kappa_B^2)\, p(\kappa_B^2|a^{\breve{\xi}},c^{\breve{\xi}})p(c^{\breve{\xi}}) \\
& \propto \quad \prod_{j=1}^d \frac{\Gamma(\frac{2c^{\breve{\xi}}+1}{2})}{\Gamma(\frac{2c^{\breve{\xi}}}{2})\left(2\pi\phi^{\breve{\xi}}\breve{\xi}_j^2\right)^{1/2}}\left(1 + \frac{\theta_j}{2\breve{\xi}_j^2\phi^{\breve{\xi}}}\right)^{-\frac{2c^{\breve{\xi}}+1}{2}} \\
& \quad \cdot \frac{1}{2^{a^{\breve{\xi}}}B(a^{\breve{\xi}},c^{\breve{\xi}})}\left(\frac{a^{\breve{\xi}}}{c^{\breve{\xi}}}\kappa_B^2\right)^{a^{\breve{\xi}}-1}\left(1 + \frac{a^{\breve{\xi}}}{2c^{\breve{\xi}}}\kappa_B^2\right)^{-(a^{\breve{\xi}}+c^{\breve{\xi}})}(2c^{\breve{\xi}})^{\alpha_{c^{\breve{\xi}}}-1}(1-2c^{\breve{\xi}})^{\beta_{c^{\breve{\xi}}}-1}.
\end{aligned}
\tag{A6}
$$

Hence, $\log q_c(c^{\breve{\xi}})$ is given by (using $\Gamma(c^{\breve{\xi}}) = \Gamma(c^{\breve{\xi}}+1)/c^{\breve{\xi}}$):

$$
\begin{aligned}
\log q_c(c^{\breve{\xi}}) \quad = \quad & d\log\Gamma(c^{\breve{\xi}}+0.5) - d\log\Gamma(c^{\breve{\xi}}+1) + \frac{d}{2}\log c^{\breve{\xi}} \\
& - (c^{\breve{\xi}}+0.5)\left(\sum_{j=1}^d \log(4c^{\breve{\xi}}\breve{\xi}_j^2 + \theta_j\kappa_B^2 a^{\breve{\xi}}) - \sum_{j=1}^d \log(4c^{\breve{\xi}}\breve{\xi}_j^2)\right) \qquad \text{(prior on } \theta_j) \\
& - \log B(a^{\breve{\xi}},c^{\breve{\xi}}) - (a^{\breve{\xi}}-1)\log c^{\breve{\xi}} - (a^{\breve{\xi}}+c^{\breve{\xi}})\log\left(1 + \frac{a^{\breve{\xi}}\kappa_B^2}{2c^{\breve{\xi}}}\right) \qquad \text{(prior on } \kappa_B^2) \\
& + (\alpha_{c^{\breve{\xi}}}-1)\log(2c^{\breve{\xi}}) + (\beta_{c^{\breve{\xi}}}-1)(1-2c^{\breve{\xi}}) \qquad \text{(prior on } c^{\breve{\xi}}) \\
& + \log c^{\breve{\xi}} + \log(0.5 - c^{\breve{\xi}}) \qquad \text{(change of variable)}
\end{aligned}
\tag{A7}
$$

In Step (e),

---

[5]   The pdf of the $\mathcal{GIG}(p,a,b)$-distribution is given by

$$
f(x) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})}x^{p-1}e^{-\frac{1}{2}(ax+b/x)},
$$

where $K_p(z)$ is the modified Bessel function.

$$
\begin{aligned}
p(\check{\kappa}_j^2|z_{-\check{\kappa}_j^2},y) \quad &\propto p(\sqrt{\theta_j}|\check{\xi}_j^2,\check{\kappa}_j^2,\phi^\xi)p(\check{\kappa}_j^2|c^\xi) \\
&\propto (\check{\kappa}_j^2)^{1/2}\exp\left\{-\frac{\check{\kappa}_j^2}{2\phi^\xi\check{\xi}_j^2}\theta_j\right\}\times(\check{\kappa}_j^2)^{c^\xi-1}\exp\left\{-\check{\kappa}_j^2\right\} \\
&= (\check{\kappa}_j^2)^{1/2+c^\xi-1}\exp\left\{-\check{\kappa}_j^2\left(\frac{\theta_j}{2\phi^\xi\check{\xi}_j^2}+1\right)\right\},
\end{aligned}
\tag{A8}
$$

which is equal to the gamma distribution given in (26).

In Step (f), $p(d_2|z_{-d_2},y)$ is equal to following gamma distribution:

$$
\begin{aligned}
p(d_2|z_{-d_2},y) \quad &\propto p(\kappa_B^2|d_2)p(d_2|a^\xi,c^\xi) \\
&\propto (d_2)^{a^\xi}\exp\left\{-d_2\kappa_B^2\right\}(d_2)^{c^\xi-1}\exp\left\{-d_2\frac{2c^\xi}{a^\xi}\right\} \\
&= (d_2)^{a^\xi+c^\xi-1}\exp\left\{-d_2\left(\kappa_B^2+\frac{2c^\xi}{a^\xi}\right)\right\},
\end{aligned}
\tag{A9}
$$

and

$$
\begin{aligned}
p(\kappa_B^2|z_{-\kappa_B^2},y) \quad &\propto \prod_{j=1}^d p(\sqrt{\theta_j}|\check{\xi}_j^2,\check{\kappa}_j^2,\phi^\xi)p(\kappa_B^2|d_2) \\
&\propto (\kappa_B^2)^{d/2}\exp\left\{-\frac{\kappa_B^2 a^\xi}{4c^\xi}\sum_{j=1}^d\frac{\check{\kappa}_j^2}{\check{\xi}_j^2}\theta_j\right\}\times(\kappa_B^2)^{a^\xi-1}\exp\left\{-d_2\kappa_B^2\right\} \\
&= (\kappa_B^2)^{d/2+a^\xi-1}\exp\left\{-\kappa_B^2\left(\frac{a^\xi}{4c^\xi}\sum_{j=1}^d\frac{\check{\kappa}_j^2}{\check{\xi}_j^2}\theta_j+d_2\right)\right\},
\end{aligned}
\tag{A10}
$$

which is equal to the gamma distribution given in (27).

For a symmetric triple gamma prior, where $a^\xi=c^\xi$, Step (b) is modified in the following way, if Step (d) is dropped:

$$
q_a(a^\xi)=p(a^\xi|z_{-a^\xi},y)\prod_{j=1}^d p(\check{\kappa}_j^2|c^\xi=a^\xi)\propto p(a^\xi|z_{-a^\xi},y)\frac{1}{\Gamma(a^\xi)^d}\left(\prod_{j=1}^d\check{\kappa}_j^2\right)^{a^\xi},
\tag{A11}
$$

where $p(a^\xi|z_{-a^\xi},y)$ is given by (A3). If Step (b) is dropped, then Step (d) is modified in the following way:

$$
q_c(c^\xi)=p(c^\xi|z_{-c^\xi},y)\prod_{j=1}^d p(\check{\xi}_j^2|a^\xi=c^\xi)\propto p(c^\xi|z_{-c^\xi},y)\frac{1}{\Gamma(c^\xi)^d}\left(\prod_{j=1}^d\check{\xi}_j^2\right)^{c^\xi},
\tag{A12}
$$

where $p(c^\xi|z_{-c^\xi},y)$ is given by (A6).

Appendix C. Posterior Paths for the Simulated Data
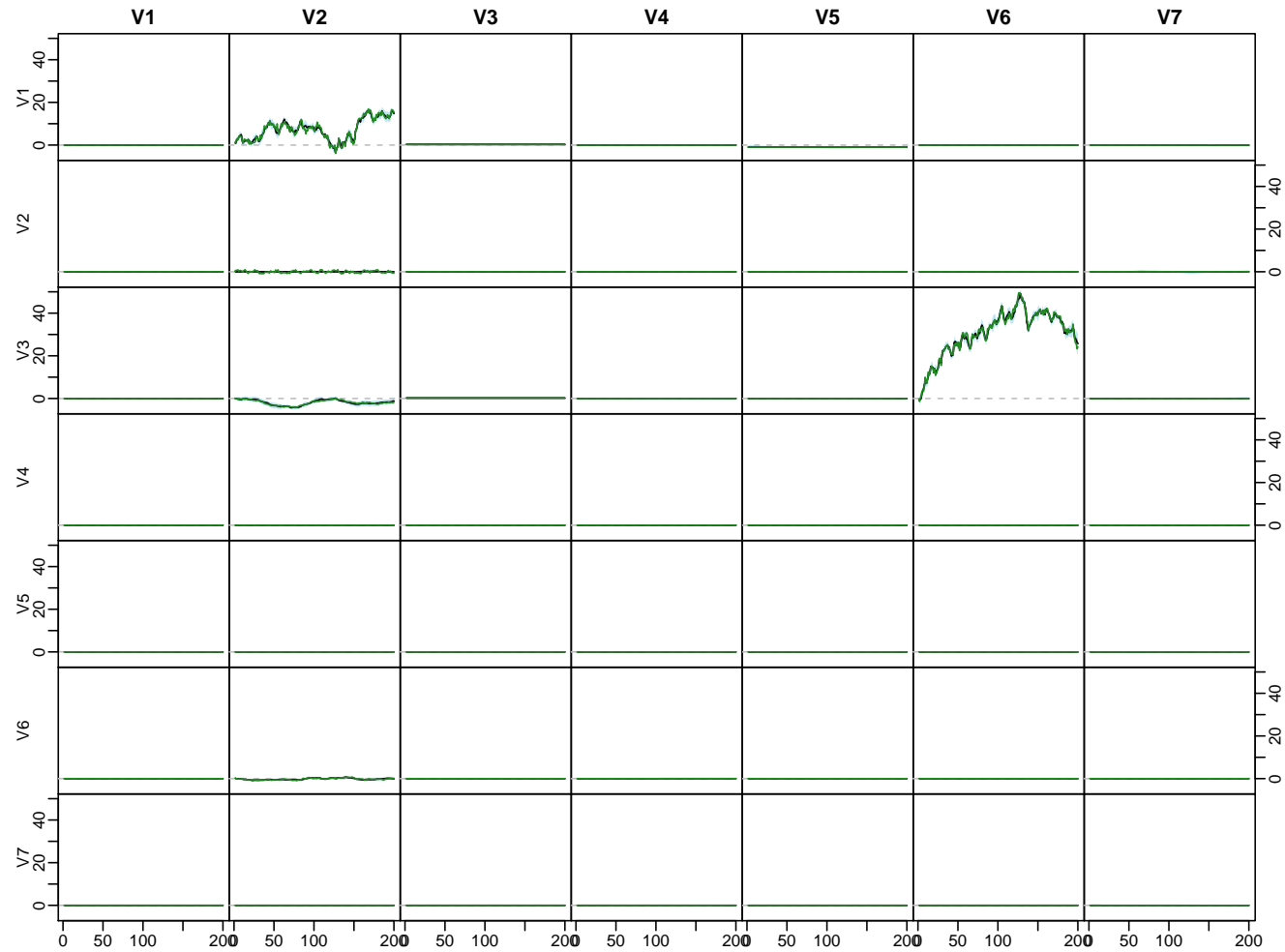

Figure A1. Each cell represents the corresponding state of the matrix $\Phi_{1,t}$, for $t = 1, \ldots, T$, for the sparse regime described in Section 5.3. The solid line is the median and the shaded areas represent 50% and 95% posterior credible intervals under the triple gamma prior.
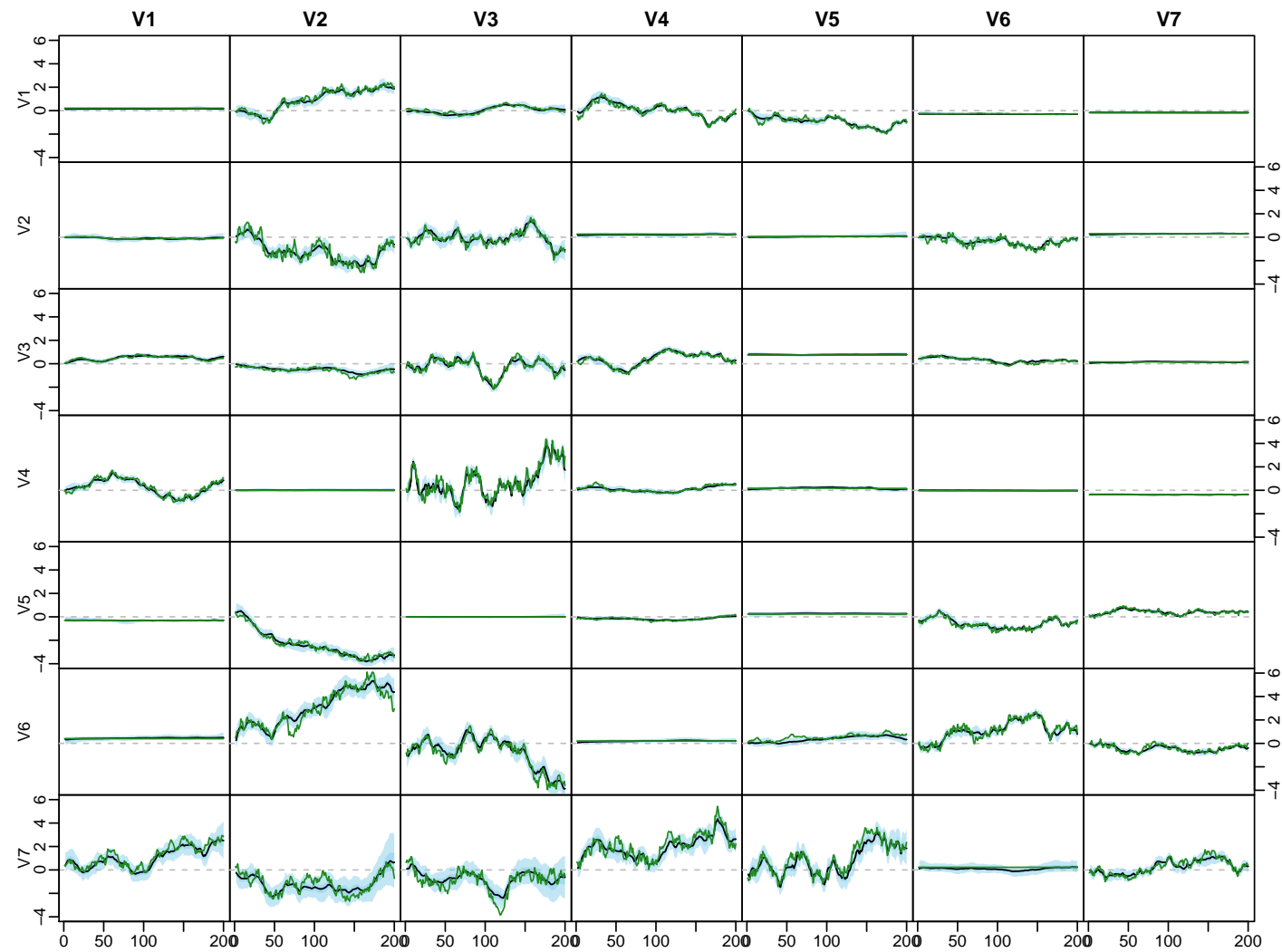
Figure A2. Each cell represents the corresponding state of the matrix $\Phi_{1,t}$, for $t = 1, \ldots, T$, for the dense regime described in Section 5.3. The solid line is the median and the shaded areas represent 50% and 95% posterior credible intervals under the triple gamma prior.

Appendix D. Application

Appendix D.1. Data Overview

Table A1. Data overview.

Variable	Abbreviation	Description	Tcode
Real output	YER	Gross domestic product (GDP) at market prices in millions of Euros, chain linked volume, calendar and seasonally adjusted data, reference year 1995.	1
Prices	YED	GDP deflator, index base year 1995. Defined as the ratio of nominal and real GDP.	1
Short-term interest rate	STN	Nominal short-term interest rate, Euribor 3-month, percent per annum	2
Investment	ITR	Gross fixed capital formation in millions of Euros, chain linked volume, calendar and seasonally adjusted data, reference year 1995.	1
Consumption	PCR	Individual consumption expenditure in millions of Euros, chain linked volume, calendar and seasonally adjusted data, reference year 1995.	1
Exchange rate	EEN	Nominal effective exchange rate, Euro area-19 countries vis-à-vis the NEER-38 group of main trading partners , index base Q1 1999.	1
Unemployment	URX	Unemployment rate, percentage of civilian work force, total across age and sex, seasonally adjusted, but not working day adjusted.	2

Note: Data was retrieved from https://eabcn.org/page/area-wide-model. Tcode = 1 indicates that differences of logs were taken, while Tcode = 2 implies that the raw data was used.
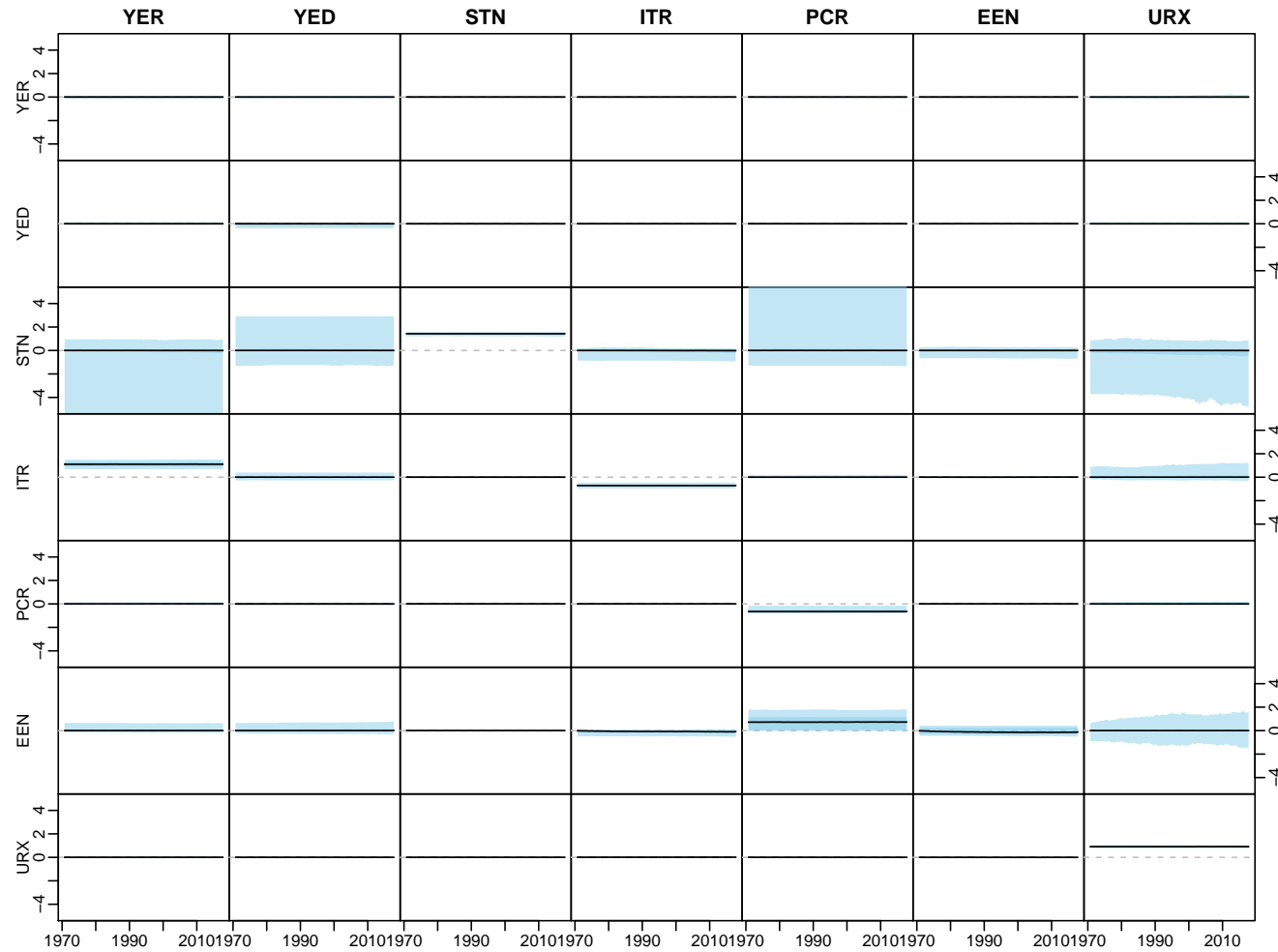
Appendix D.2. Posterior Paths

Figure A3. Each cell represents the corresponding state of the matrix $\Phi_{1,t}$, for $t = 1, \ldots, T$, for the data described in Section 5.4. The solid line is the median and the shaded areas represent 50% and 95% posterior credible intervals under the triple gamma prior.
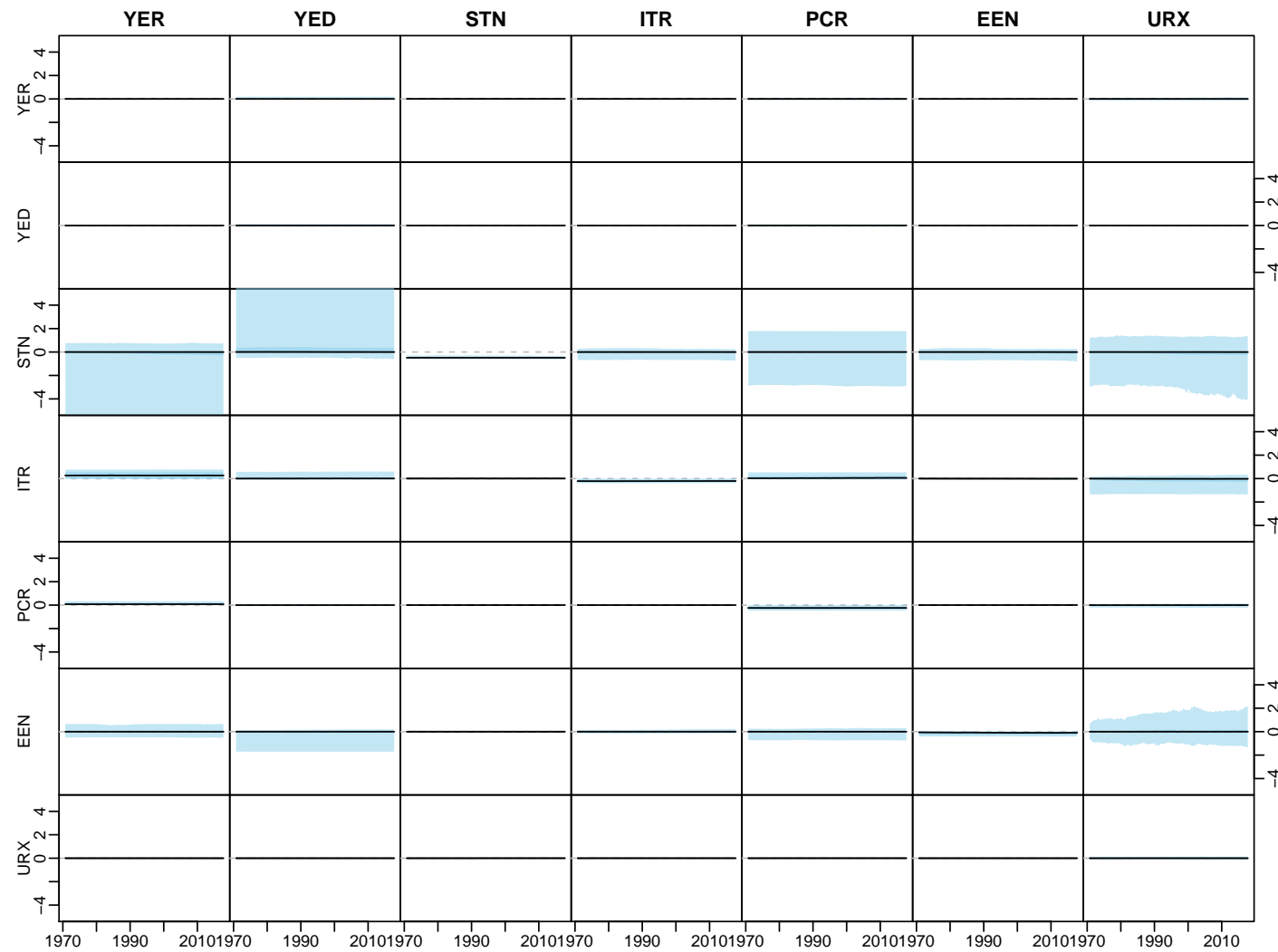
Figure A4. Each cell represents the corresponding state of the matrix $\mathbf{\Phi}_{2,t}$, for $t = 1, \ldots, T$, for the data described in Section 5.4. The solid line is the median and the shaded areas represent 50% and 95% posterior credible intervals under the triple gamma prior.

References

Abramowitz, Milton, and Irene A. Stegun, eds. 1973. *Handbook of Mathematical Functions*. New York: Dover Publications.

Armagan, Artin, David B. Dunson, and Merlise Clyde. 2011. Generalized beta mixtures of Gaussians. In *Advances in Neural Information Processing Systems*. Vancouver: NIPS, pp. 523–31.

Belmonte, Miguel, Gary Koop, and Dimitris Korobolis. 2014. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting* 33: 80–94. [CrossRef]

Berger, James O. 1980. A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *The Annals of Statistics* 8: 716–61. [CrossRef]

Bhadra, Anindya, Jyotishka Datta, Nicholas G. Polson, and Brandon Willard. 2017a. The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis* 12: 1105–31. [CrossRef]

Bhadra, Anindya, Jyotishka Datta, Nicholas G. Polson, and Brandon Willard. 2017b. Horseshoe regularization for feature subset selection. *arXiv*: arXiv:1702.07400.

Bhadra, Anindya, Jyotishka Datta, Nicholas G. Polson, and Brandon Willard. 2019. Lasso meets horseshoe: A survey. *Statistical Science* 34: 405–27. [CrossRef]

Bitto, Angela, and Sylvia Frühwirth-Schnatter. 2019. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210: 75–97. [CrossRef]

Brown, Philip J., Marina Vannucci, and Tom Fearn. 2002. Bayes model averaging with selection of regressors. *Journal of the Royal Statistical Society, Ser. B* 64: 519–36. [CrossRef]

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2019. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics* 212: 137–54. [CrossRef]

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott. 2009. Handling sparsity via the horseshoe. *Journal of Machine Learing Research W&CP* 5: 73–80.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott. 2010. The horseshoe estimator for sparse signals. *Biometrika* 97: 465–80. [CrossRef]

Chan, Joshua C. C., and Eric Eisenstat. 2016. Bayesian model comparison for time-varying parameter VARs with stochastic volatilty. *Journal of Applied Econometrics* 218: 1–24.

Cottet, Remy, Robert J. Kohn, and David J. Nott. 2008. Variable selection and model averaging in semiparametric overdispersed generalized linear models. *Journal of the American Statistical Association* 103: 661–71. [CrossRef]

Del Negro, Marco, and Giorgio E. Primiceri. 2015. Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum. *The Review of Economic Studies* 82: 1342–45. [CrossRef]

Eisenstat, Eric, Joshua C.C. Chan, and Rodney W. Strachan. 2014. Stochastic model specification search for time-varying parameter VARs. *SSRN Electronic Journal 01/2014*. [CrossRef]

Fagan, Gabriel, Jerome Henry, and Ricardo Mestre. 2005. An area-wide model for the euro area. *Economic Modelling* 22: 39–59. [CrossRef]

Fahrmeir, Ludwig, Thomas Kneib, and Susanne Konrath. 2010. Bayesian regularisation in structured additive regression: A unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing* 20: 203–19. [CrossRef]

Feldkircher, Martin, Florian Huber, and Gregor Kastner. 2017. Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian VARs. *arXiv*: arXiv:1711.00564.

Fernández, Carmen, Eduardo Ley, and Mark F. J. Steel. 2001. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100: 381–427. [CrossRef]

Figueiredo, Mario A. T. 2003. Adaptive sparseness for supervised learning. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 25: 1150–59. [CrossRef]

Frühwirth-Schnatter, Sylvia. 2004. Efficient Bayesian parameter estimation. In *State Space and Unobserved Component Models: Theory and Applications*. Edited by Andrew Harvey, Siem Jan Koopman and Neil Shephard. Cambridge: Cambridge University Press, pp. 123–51.

Frühwirth-Schnatter, Sylvia, and Regina Tüchler. 2008. Bayesian parsimonious covariance estimation for hierarchical linear mixed models. *Statistics and Computing* 18: 1–13. [CrossRef]

Frühwirth-Schnatter, Sylvia, and Helga Wagner. 2010. Stochastic model specification search for Gaussian and partially non-Gaussian state space models. *Journal of Econometrics* 154: 85–100. [CrossRef]

Gelman, Andrew. 2006. Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis* 1: 515–34. [CrossRef]

Griffin, Jim E., and Phil J. Brown. 2011. Bayesian hyper-lassos with non-convex penalization. *Australian & New Zealand Journal of Statistics* 53: 423–42.

Griffin, Jim E., and Phil J. Brown. 2017. Hierarchical shrinkage priors for regression models. *Bayesian Analysis* 12: 135–59. [CrossRef]

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi. 1994. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* 12: 371–417.

Johnstone, Iain M., and Bernard W. Silverman. 2004. Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *The Annals of Statistics* 32: 1594–649. [CrossRef]

Kalli, Maria, and Jim E. Griffin. 2014. Time-varying sparsity in dynamic regression models. *Journal of Econometrics* 178: 779–93. [CrossRef]

Kastner, Gregor. 2016. Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software* 69: 1–30. [CrossRef]

Kastner, Gregor, and Sylvia Frühwirth-Schnatter. 2014. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics and Data Analysis* 76: 408–23. [CrossRef]

Kleijn, Richard, and Herman K. van Dijk. 2006. Bayes model averaging of cyclical decompositions in economic time series. *Journal of Applied Econometrics* 21: 191–212. [CrossRef]

Koop, Gary, and Dimitris Korobilis. 2013. Large time-varying parameter VARs. *Journal of Econometrics* 177: 185–98. [CrossRef]

Koop, Gary, and Simon M. Potter. 2004. Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal* 7: 550–65. [CrossRef]

Kowal, Daniel R., David S. Matteson, and David Ruppert. 2019. Dynamic shrinkage processes. *Journal of the Royal Statistical Society, Ser. B* 81: 673–806. [CrossRef]

Ley, Eduardo, and Mark F. J. Steel. 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics* 24: 651–74. [CrossRef]

Makalic, Enes, and Daniel F. Schmidt. 2016. A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters* 23: 179–82. [CrossRef]

Nakajima, Jouchi. 2011. Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies* 29: 107–42.

Park, Trevor, and George Casella. 2008. The Bayesian Lasso. *Journal of the American Statistical Association* 103: 681–86. [CrossRef]

Pérez, Maria-Eglée, Luis Raúl Pericchi, and Isabel Cristina Ramírez. 2017. The scaled beta2 distribution as a robust prior for scales. *Bayesian Analysis* 12: 615–37. [CrossRef]

Polson, Nicholas G., and James G. Scott. 2011. Shrink globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian Statistics 9*. Edited by José M. Bernardo, M. J. Bayarri, James O. Berger, Phil Dawid, David Heckerman, Adrian F. M. Smith and Mike West. Oxford: Oxford University Press, pp. 501–38.

Polson, Nicholas G., and James G. Scott. 2012a. Local shrinkage rules, Lévy processes, and regularized regression. *Journal of the Royal Statistical Society, Ser. B* 74: 287–311. [CrossRef]

Polson, Nicholas G., and James G. Scott. 2012b. On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* 7: 887–902. [CrossRef]

Primiceri, Giorgio E. 2005. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* 72: 821–52. [CrossRef]

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92: 179–91. [CrossRef]

Ročková, Veronika, and Kenichiro McAlinn. 2020. Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis*.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller. 2004. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *The American Economic Review* 94: 813–35. [CrossRef]

Scheipl, Fabian, and Thomas Kneib. 2009. Locally adaptive Bayesian p-splines with a normal-exponential-gamma prior. *Computational Statistics and Data Analysis* 53: 3533–52. [CrossRef]

Strawderman, William E. 1971. Proper Bayes minimax estimators of the multivariate normal mean. *The Annals of Statistics* 42: 385–88. [CrossRef]

Tibshirani, Ryan. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Ser. B* 58: 267–88. [CrossRef]

van der Pas, Stéphanie, Bas Kleijn, and Aad van der Vaart. 2014. The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics* 8: 2585–618. [CrossRef]

Zhang, Yan, Brian J. Reich, and Howard D. Bondell. 2017. High dimensional linear regression via the R2-D2 shrinkage prior. Technical report. *arXiv*: arXiv:1609.00046v2.