Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Loss-based prior for the degrees of freedom of the Wishart distribution

Luca Rossini [a,b,*], Cristiano Villa [c], Sotiris Prevenas [d], Rachel McCrea [e]

[a] Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio 7, Milan, 20122, Italy
[b] Fondazione Eni Enrico Mattei, Corso Magenta 63, Milan, 20123, Italy
[c] Division of Natural Sciences, Duke-Kunshan University, Duke Avenue 8, Kunshan, 215316, China
[d] School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7ND, United Kingdom
[e] Department of Mathematics and Statistics, Lancaster University, Fylde College, Lancaster, LA1 4YF, United Kingdom

ARTICLE INFO

Article history:
Received 7 April 2023
Revised 5 April 2024
Accepted 5 April 2024
Available online xxx

Keywords:
Forecasting
Global-local shrinkage prior
Loss-based prior
Macroeconomic data
Vector autoregressive models

ABSTRACT

Motivated by the proliferation of extensive macroeconomic and health datasets necessitating accurate forecasts, a novel approach is introduced to address Vector Autoregressive (VAR) models. This approach employs the global-local shrinkage-Wishart prior. Unlike conventional VAR models, where degrees of freedom are predetermined to be equivalent to the size of the variable plus one or equal to zero, the proposed method integrates a hyperprior for the degrees of freedom to account for the uncertainty in the parameter values. Specifically, a loss-based prior is derived to leverage information regarding the data-inherent degrees of freedom. The efficacy of the proposed prior is demonstrated in a multivariate setting both for forecasting macroeconomic data, and Dengue infection data.

1. Introduction

Macroeconomic and health forecasting holds pivotal importance. Advancements in modeling techniques and the availability of large datasets have enhanced the forecasting of macroeconomic and financial variables (Stock and Watson, 2002; De Mol et al., 2008; Bańbura et al., 2010; Koop and Korobilis, 2013; Huber and Feldkircher, 2019). Extracting relevant information from the data regarding the processes of interest is beneficial. It is demonstrated in leveraging intrinsic information within variables of interest to provide an adaptable representation of prior uncertainty.

The Vector Autoregressive (VAR) model, introduced by Sims (1980), is widely used. It is employed for representing the relationships between quantities of interest, given its flexibility to accommodate multiple time series. Within the Bayesian framework, substantial research has been carried out to avoid over parametrization and overfitting (e.g. Doan et al., 1984; Litterman, 1986; Sims and Zha, 1998), mainly through the definition of suitable shrinkage prior distributions, or by employing appropriate hierarchical structures (e.g. Bańbura et al., 2010; Carriero et al., 2015; 2019; Huber and Feldkircher, 2019).

We focus on the global-local shrinkage prior for the matrix of coefficients relying on the Horseshoe prior distribution (Carvalho et al., 2010). The Horseshoe prior (Follett and Yu, 2019) was used on the reduced form of the VAR matrix of coef-

* Corresponding author.
E-mail address: luca.rossini@unimi.it (L. Rossini).

https://doi.org/10.1016/j.ecosta.2024.04.001

ficients by Bernardi et al. (2024) and Gruber and Kastner (2023) to shrink the coefficients towards zeros. For the covariance matrix, general approaches fix the degrees of freedom of the Wishart distribution to be equal to the size of the variable plus one (Koop and Korobilis, 2010) or equal to zero (Uhlig, 2005). As stated in Koop and Korobilis (2010), this choice is a trade-off between avoiding overfitting and allowing for flexibility. We derive a hyperprior for the degrees of freedom based on the loss-based priors introduced in Villa and Walker (2015). This hyperprior accounts for the uncertainty of the parameter values and exploits the information about the number of degrees of freedom contained in the data.

The loss-based hyperprior is based on the measure of the loss that would be incurred if the wrong model was chosen and depends solely on the probability distribution used to model the data. Hence no value is set for the degrees of freedom, and it impacts the posterior distribution and model forecasts. The performance of the hyperprior is assessed using simulation and contrasted with alternative approaches, such as a flat prior (Uhlig, 2005).

The merits of the hyperprior are demonstrated by forecasting different macroeconomic variables from the Federal Reserve Economic Data (FRED) dataset (McCracken and Ng, 2016; 2020). The proposed prior has two important advantages compared to fixing the degrees of freedom. Firstly, in the absence of any prior information about the true variance-covariance matrix, the approach is more robust since the process is combined with information in the data. Secondly, the degrees of freedom are evaluated through time using a rolling window procedure, where they increase after 2000 and decrease during the 2009 financial crisis.

Additionally, we analyse the Google Dengue Trends (GDT) dataset for ten countries, which is a query-based reporting system for infectious disease (Carneiro and Mylonakis, 2009; Strauss et al., 2017). In particular, dengue is a viral infection transmitted by mosquitoes, which is present in South American and Asian countries. The forecasting results demonstrate the flexibility of the proposed approach.

The paper is structured as follows: Section 2 describes the VAR model and derives the proposed loss-based prior for the degrees of freedom of the Wishart distribution. In Section 3, using simulation, we compare the proposed hyperprior to the model that assumes fixed degrees of freedom. Section 4 presents the forecasting of macroeconomic variables, and the forecasting of the Google Dengue Trends data are presented in Section 5. Section 6 concludes the paper.

2. VAR model for forecasting and the novel loss-based prior

Let $\mathbf{y}_t$ be the m-dimensional vector of observations, for $t = 1, \dots, T$. A VAR model with p lags is given by

$$\mathbf{y}_t = \sum_{j=1}^{p} A_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where A_j is a $(m \times m)$ matrix of coefficients, and $\boldsymbol{\varepsilon}_t$ is an m-dimensional vector of independent and identically normally distributed error terms centered on 0, and with covariance matrix Σ, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \Sigma)$.

Eq. (1) can be written concisely as

$$Y = XA + E,$$

where Y is a $(T \times m)$ matrix constructed as $Y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)'$, and $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)'$ is a $(T \times k)$ matrix containing the lagged response variables, where $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \dots, \mathbf{y}'_{t-p})$. $A = (A_1, A_2, \dots, A_p)$ is a $(k \times m)$ matrix of coefficients and $E = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T)'$ is a $(T \times m)$ matrix of errors. In a vectorized form, the VAR model of order p is defined as

$$\mathbf{y} = (I_m \otimes X)\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = \text{vec}(Y)$, $\boldsymbol{\alpha} = \text{vec}(A)$ and $\boldsymbol{\varepsilon} = \text{vec}(E)$ with distribution $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \Sigma \otimes I_T)$ and $\otimes$ is the Kronecker product.

We adopt a global-local shrinkage-Wishart prior for the parameters of the model. We assume a Horseshoe prior distribution (Carvalho et al., 2010) for the matrix of coefficients and a Wishart distribution for the precision matrix $\Sigma^{-1} \sim \mathcal{W}(\underline{\nu}, \underline{S}^{-1})$. We set the hyperparameters for the Wishart prior equal to $\underline{\nu} = m + 1$ and $\underline{S} = I_m$. As a robustness check, we ran a Normal-Wishart prior with hyperparameters for the Normal prior equal to $\underline{\boldsymbol{\alpha}} = \mathbf{0}$ and $\underline{V} = 10 \cdot I_{mk}$ (the results are available upon request).

The general Horseshoe prior for each element of the vectorized matrix of coefficient, $\boldsymbol{\alpha}$, takes the form

$$\begin{aligned} \alpha_j | (\lambda_j^\alpha)^2, (\tau^\alpha)^2 &\sim \mathcal{N}(0, (\lambda_j^\alpha)^2 (\tau^\alpha)^2), \\ \lambda_j^\alpha &\sim \mathcal{C}^+(0, 1), \\ \tau^\alpha &\sim \mathcal{C}^+(0, 1), \end{aligned}$$

where $\mathcal{C}^+(\cdot, \cdot)$ denotes the half-Cauchy distribution, λ_j^α is the local shrinkage parameter, and τ^α is the global shrinkage parameter, for $j = 1, \dots, k \cdot m$. For the posterior distribution of α, and the global and local shrinkage parameters λ_j^α and τ^α, we refer to Cross et al. (2020) and the algorithm proposed by Makalic and Schmidt (2016).

2.1. Loss-based hyperprior

The usual assumption on the degrees of freedom of the Wishart distribution is to set the parameter to $\nu = m + 1$. In this section, we derive the loss-based prior distribution for ν. The parameter is assumed discrete, and the objective method introduced in Villa and Walker (2015) is used to construct the prior.

Consider a Bayesian model with sampling distribution $f(x|\theta)$, characterized by the discrete parameter $\theta \in \Theta$, and prior $\pi(\theta)$. A mass is assigned to each value of the parameter that is proportional to the Kullback–Leibler divergence between the model defined by θ and the nearest one. Let us define $f(x|\theta)$ as the true model not chosen and $f(x|\theta')$ denotes the nearest model to the true model. The Kullback-Leibler divergence between $f(x|\theta)$ and $f(x|\theta')$ represents the loss in information The prior on θ is

$$\pi(\theta) \propto \exp\left\{\min_{\theta' \neq \theta \in \Theta} KL\big(f(x|\theta)\|f(x|\theta')\big)\right\} - 1, \tag{2}$$

where $KL\big(f(x|\theta)\|f(x|\theta')\big)$ represents the Kullback–Leibler divergence between the models. A more detailed derivation of the prior in Eq. (2) is illustrated in Appendix A.

To derive the prior for ν, the Kullback–Leibler divergence between two Wishart distributions that share the same scale matrix V and differ in the number of degrees of freedom, say W_ν and $W_{\nu+c}$, is required. The probability density function of a Wishart distribution with parameters V and ν is given by:

$$W(X|V,\nu) = \frac{|X|^{(\nu-m-1)/2}\exp(-\operatorname{Tr}(V^{-1}X)/2)}{2^{\frac{\nu m}{2}}|V|^{\nu/2}\Gamma_m(\nu/2)},$$

where $\Gamma_m(\cdot)$ is the multivariate Gamma function, and $\operatorname{Tr}(\cdot)$ is the trace function. Thus the Kullback–Leibler divergence between two Wishart distributions is given by:

$$KL(W_\nu\|W_{\nu+c}) = \log\left\{\frac{\Gamma_m\left(\frac{\nu+c}{2}\right)}{\Gamma_m\left(\frac{\nu}{2}\right)}\right\} - \frac{c}{2}\psi_m\left(\frac{\nu}{2}\right),$$

where $\psi_m(\cdot)$ is the multivariate digamma function defined as $\psi_m(x) = \sum_{i=1}^m \psi(x+(1-i)/2)$, $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function, and $c \in \mathbb{Z}$. As $KL(W_\nu\|W_{\nu+c})$ is a convex function of c, and its global minimum is at $c=0$, the nearest Wishart distribution to W_ν will be $W_{\nu+c}$ for either $c=-1$ or $c=1$. Theorem 1 shows that the Kullback–Leibler divergence between W_ν and $W_{\nu+c}$ is minimized for $c=1$.

Theorem 1. *Consider two Wishart distributions, W_ν and $W_{\nu+c}$, with the same scale matrix and ν and $\nu+c$ degrees of freedom, respectively, and $c \neq 0$ is an integer. Then, the Kullback–Leibler divergence between W_ν and $W_{\nu+c}$ is minimum for $c=1$.*

Proof. For $c=1$, the Kullback–Leibler divergence is

$$KL(W_\nu\|W_{\nu+1}) = \log\Gamma_m\left(\frac{\nu+1}{2}\right) - \log\Gamma_m\left(\frac{\nu}{2}\right) - \frac{1}{2}\psi_m\left(\frac{\nu}{2}\right),$$

while for $c=-1$, we obtain

$$KL(W_\nu\|W_{\nu-1}) = \log\Gamma_m\left(\frac{\nu-1}{2}\right) - \log\Gamma_m\left(\frac{\nu}{2}\right) + \frac{1}{2}\psi_m\left(\frac{\nu}{2}\right).$$

Taking the difference of the two divergences:

$$KL(W_\nu\|W_{\nu+1}) - KL(W_\nu\|W_{\nu-1}) = \log\Gamma_m\left(\frac{\nu+1}{2}\right) - \log\Gamma_m\left(\frac{\nu-1}{2}\right) - \psi_m\left(\frac{\nu}{2}\right) \tag{3}$$

$$= \log\left[\frac{\Gamma(\nu)}{2^m\Gamma(\nu-m)}\right] - \psi_m\left(\frac{\nu}{2}\right). \tag{4}$$

We will prove that Eq. (4) is always negative for any ν, m such that $\nu > m \geq 2$. As $\nu > m$, $\nu = m+k$, for $k = 1, 2, \ldots$, and $m \geq 2$, thus we obtain that the minimum Kullback–Leibler divergence is achieved at $c=1$ if

$$\log\left\{\frac{\Gamma(m+k)}{2^m\Gamma(k)}\right\} < \psi_m\left(\frac{m+k}{2}\right). \tag{5}$$

To prove the inequality in Eq. (5), we rely on

$$\psi_{m+1}\left(\frac{\nu+1}{2}\right) = \psi_m\left(\frac{\nu}{2}\right) + \psi\left(\frac{\nu+1}{2}\right), \tag{6}$$

and

$$\log\left(x-\tfrac{1}{2}\right) < \psi(x), \tag{7}$$

where Eq. (6) comes from the definition of the multivariate digamma function, and inequality (7) can be deduced from $\log(x+\frac{1}{2}) - \frac{1}{x} < \psi(x) < \log(x+e^{-\gamma}) - \frac{1}{x}$, (Elezovic et al., 2000) for $x > \frac{1}{2}$ and with γ equal to 0.57721 as the Euler-Mascheroni constant.

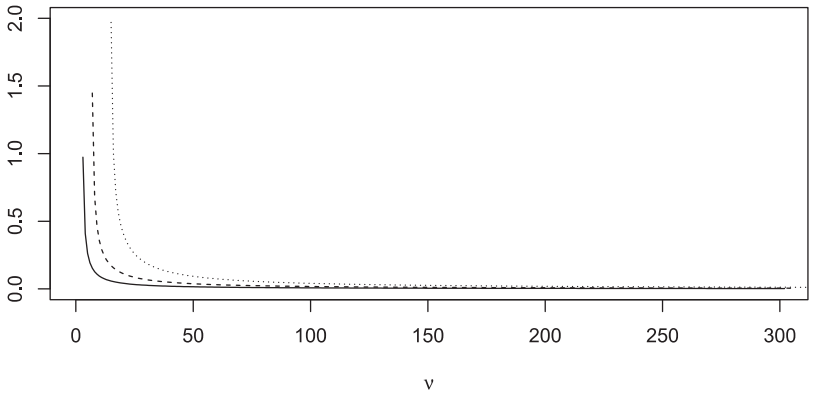

Fig. 1. Loss-based prior distribution (unnormalised) for ν in dimensionality $m = 3$ (continuous line), $m = 7$ (dashed line), and $m = 15$ (dotted line).

Assuming the inequality holds for m, considering $m + 1$:

$$\log\left\{\frac{\Gamma(m+1+k)}{2^{m+1}\Gamma(k)}\right\} < \psi_{m+1}\left(\frac{m+1+k}{2}\right)$$
$$\log\left\{\frac{\Gamma(m+k)(m+k)}{2^{m}\Gamma(k)2}\right\} < \psi_{m}\left(\frac{m+k}{2}\right) + \psi\left(\frac{m+1+k}{2}\right)$$
$$\log\left\{\frac{\Gamma(m+k)}{2^{m}\Gamma(k)}\right\} + \log\left\{\frac{m+k}{2}\right\} < \psi_{m}\left(\frac{m+k}{2}\right) + \psi\left(\frac{m+1+k}{2}\right).$$

Within the last inequality, we have $\log\frac{m+k}{2} < \psi\left(\frac{m+1+k}{2}\right)$ as a consequence of the result in inequality (7). Thus, if the inequality (5) holds for m, then it holds for $m + 1$, and it holds for any k. The smallest possible value m is $m = 2$, as

$$\log\left\{\frac{\Gamma(2+k)}{2^{2}\Gamma(k)}\right\} < \psi_{2}\left(\frac{2+k}{2}\right)$$
$$\log\left\{\frac{(k+1)k}{2^{2}}\right\} < \psi\left(\frac{2+k}{2}\right) + \psi\left(\frac{1+k}{2}\right),$$

where $\log\left(\frac{k}{2}\right) < \psi\left(\frac{k+1}{2}\right)$ and $\log\left(\frac{k+1}{2}\right) < \psi\left(\frac{k+2}{2}\right)$ due to inequality (7). Thus, inequality (5) holds for $m = 2$ and, subsequently, it holds for any m. □

We can define the objective prior distribution for ν as

$$\pi(\nu) \propto \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu+1-m}{2}\right)} e^{-\frac{1}{2}\sum_{i=1}^{m}\psi\left(\frac{\nu+1-i}{2}\right)} - 1. \tag{8}$$

Figure 1 shows the loss-based prior for ν for three different values of $m \in \{3, 7, 15\}$, which represents the dimensionality of the macroeconomic dataset that we analyse in Section 4.1. Each line yields similar patterns and shows that the loss-based prior probability distribution decreases as the degrees of freedom ν increases.

2.2. Properness of the posterior for ν

Let us assume that we observe one random matrix Σ^{-1} from the Wishart $W(S_0^{-1}, \nu)$, then, the likelihood function is given by

$$p(y|\alpha, \Sigma) = (2\pi)^{-\frac{mT}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\left[y - (I_m \otimes X)\alpha\right]'\left(\Sigma^{-1} \otimes I_T\right)\left[y - (I_m \otimes X)\alpha\right]\right\}$$
$$\propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left[(Y - XA)'(Y - XA)\Sigma^{-1}\right]\right\}. \tag{9}$$

Using the loss-based prior for ν in Eq. (2), we obtain the posterior distribution for the number of degrees of freedom as

$$p(\nu|\Sigma^{-1}) \propto \pi(\nu)\pi\left(\Sigma^{-1}|\nu, S_0^{-1}\right)$$

$$\propto \left\{ \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu+1-m}{2}\right)} e^{-\frac{1}{2}\sum_{i=1}^{m}\psi\left(\frac{\nu+1-i}{2}\right)} - 1 \right\} \left\{ \frac{|\Sigma^{-1}|^{(\nu-m-1)/2} e^{\left(-\mathrm{Tr}(S_0\Sigma^{-1})/2\right)}}{2^{\frac{\nu m}{2}} |S_0^{-1}|^{\nu/2} \Gamma_m\left(\frac{\nu}{2}\right)} \right\}. \tag{10}$$

Theorem 2 shows that the marginal posterior distribution for ν is proper.

Theorem 2. *The posterior distribution for the number of degrees of freedom ν in Eq. (10) is proper.*

Proof. We prove by using Abel's test of convergence that $\sum_{\nu=m}^{\infty} p(\nu|\Sigma^{-1}) < \infty$. The sequence $\{\pi(\nu)\}$ is bounded since $\pi(m) > \pi(\nu) > 0$ and is also monotone (decreasing). We show that $\sum_{\nu=m}^{\infty} \pi(\Sigma^{-1}|\nu, S_0^{-1}) < \infty$ by using the ratio test

$$\begin{aligned} R_\nu &= \frac{|\Sigma^{-1}|^{(\nu-m)/2}}{2^{\frac{(\nu+1)m}{2}} |S_0^{-1}|^{(\nu+1)/2} \Gamma_m\left(\frac{\nu+1}{2}\right)} \times \frac{2^{\frac{\nu m}{2}} |S_0^{-1}|^{\nu/2} \Gamma_m\left(\frac{\nu}{2}\right)}{|\Sigma^{-1}|^{(\nu-m-1)/2}} \\ &= \frac{|\Sigma^{-1}|^{1/2} \Gamma_m\left(\frac{\nu}{2}\right)}{2^{m/2} |S_0^{-1}|^{1/2} \Gamma_m\left(\frac{\nu+1}{2}\right)} \\ &= \frac{|\Sigma^{-1}|^{1/2} \Gamma\left(\frac{\nu+1-m}{2}\right)}{2^{m/2} |S_0^{-1}|^{1/2} \Gamma\left(\frac{\nu+1}{2}\right)}. \end{aligned}$$

Thus, we have

$$\lim_{\nu\to\infty} \left\{ \frac{\Gamma\left(\frac{\nu+1-m}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} \right\} = 0,$$

so $\lim_{\nu\to\infty} R_\nu = 0$ (Abramowitz and Stegun, 1972), therefore, the posterior distribution for ν is a proper distribution. □

2.3. Gibbs sampling algorithm

Based on the proposed loss-based prior distribution, we provide a Gibbs sampler algorithm. The algorithm follows the typical steps observed in multivariate time series analysis (see, Koop and Korobilis, 2010; Follett and Yu, 2019), where the new step relates to the full conditional posterior distribution of ν. We implement a Metropolis-Hastings algorithm since the posterior distribution of ν is not available in closed form.

To summarize, the Gibbs sampler is based on the following steps:

(i) Update the vectorized matrix of coefficients, $\boldsymbol{\alpha}$ given the data $\mathbf{y}$ and Σ^{-1} by using the corrected triangular algorithm of Carriero et al. (2022).
(ii) Update the precision matrix Σ^{-1} given $\boldsymbol{\alpha}, \mathbf{y}$ and ν from a Wishart distribution.
(iii) Update the local and global shrinkage parameters λ_j^{α} and τ^{α} given the vectorized matrix of coefficients, $\boldsymbol{\alpha}$, as in Makalic and Schmidt (2016).
(iv) Update the degree of freedom ν given Σ^{-1} by using a Metropolis-Hastings algorithm with a symmetric random walk proposal.

For the model with fixed ν equal to 0 or $m+1$, the Gibbs sampler is based only on Steps (i)–(iii).

3. Simulation study

We compare the performance of our loss-based hyperprior using different simulation scenarios to the case of fixed ν equal to 0 or $m+1$. We generate the data from a VAR model with one lag, where the elements of the matrix of coefficients are drawn from a $U(-0.95, 0.95)$ distribution and then stationarity conditions are checked. We consider small, medium, and large numbers of response variables, where m is equal to 5, 10, and 20, respectively, and time, T, is equal to 30 or 100. We have considered different combinations for the choice of the degrees of freedom when generating the data: for each dimension m, we have chosen ν equal to $\{5, 10, 15\}$ for $m = 5$, $\{10, 15, 20\}$ for $m = 10$, and $\{20, 24, 26\}$ for $m = 20$. In Appendix B, we generate data as in the macroeconomic application with a time dimension equal to 240 and the number of response variables equal to 3, 7, and 15. The choice of ν is $\{3, 5, 7\}$ for $m = 3$; $\{7, 10, 13\}$ for $m = 7$, and $\{15, 20, 25\}$ for $m = 15$.

For the comparative approaches, we use a flat prior as in Uhlig (2005), or we treat the degrees of freedom as fixed at $\nu = m+1$. We assume an identity matrix for the prior scale matrix of the Wishart distribution and a Horseshoe prior for the matrix of coefficients since the interest of our simulation experiment is in the evaluation of the covariance matrix.

For each dataset, and each posterior sample, we estimate the posterior means of the vectorized matrix of coefficients and of the covariance matrix Σ. We compute the Root Mean Absolute Deviation (RMAD) between the posterior means and the true parameter values as

$$RMAD = \left[\frac{1}{N} \sum_{i=1}^{N} |\theta - \hat{\theta}| \right]^{\frac{1}{2}},$$

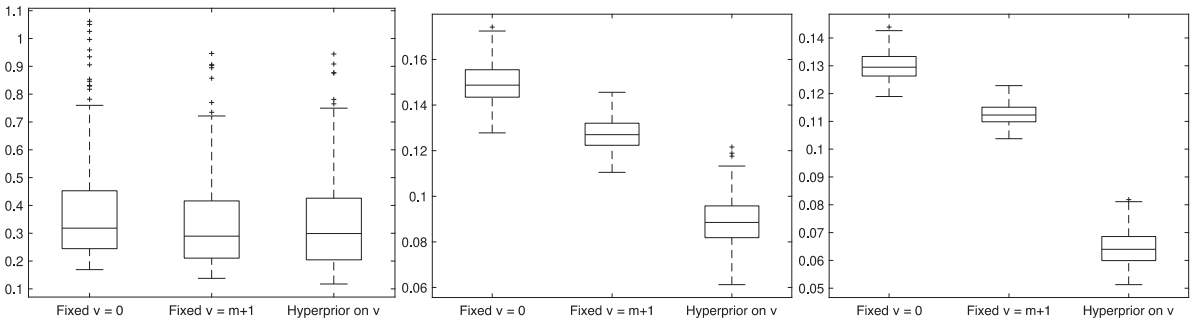

Fig. 2. Monte Carlo Simulation - Root Mean Absolute Deviations (RMAD) of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 5$ and sample size $T = 30$. Results are reported for data generated from a Wishart distribution with $\nu = 5$ (left), $\nu = 10$ (center), and $\nu = 15$ (right).

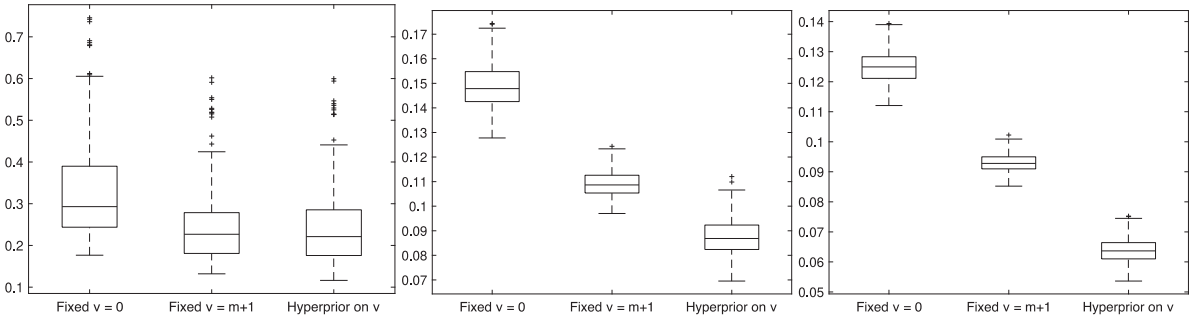

Fig. 3. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 10$ and sample size $T = 30$. Results are reported for data generated from a Wishart distribution with $\nu = 10$ (left), $\nu = 15$ (center), and $\nu = 20$ (right).

Table 1
Monte Carlo Simulation - RMAD of the impulse response functions (IRF) for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 5$ and sample size $T = 30$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 5$			$\nu = 10$			$\nu = 15$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.570	0.995	0.994	0.337	0.999	0.995	0.303	1.00	0.998
3	0.640	1.007	1.007	0.443	0.978	0.932	0.489	0.978	0.888
5	0.726	0.971	0.961	0.935	0.935	0.791	1.320	0.944	0.730
7	1.278	0.915	0.895	2.422	0.896	0.691	4.08	0.917	0.615

where θ is the matrix of coefficients or the covariance matrix. The number of parameters estimated, N is equal to m^2 for the covariance matrix and depends on the lags for the matrix of coefficients.

The Gibbs sampler is run for 6000 iterations with a burn-in of 1000 iterations and repeated 250 times. For each scenario, we present three boxplots of RMAD corresponding to degrees of freedom being fixed to 0 and $m + 1$, and the loss-based prior. Figure 2 shows the results for the RMAD for the case with $m = 5$ and $T = 30$ when the data are generated with $\nu = 5$ (left panel), with $\nu = 10$ (center), and with $\nu = 15$ (right).

From Figure 2, we observe that differences in RMAD between the alternative priors increase when increasing ν. The left panel shows no difference between the three cases, except for the outliers, which are smaller for our proposed hyperprior. Increasing ν to 10 and 15 leads to smaller RMAD when our hyperprior is used.

As a second measure of evaluation of the proposed prior, in Table 1 we provide the RMAD for the orthogonal (Choleski) impulse response function evaluated at four different horizons ($h = 1, 3, 5$ and 7). In particular, the column called "Fixed $\nu = 0$" refers to the Uhlig (2005) flat prior and it is considered as the benchmark, and the other two columns (called "Fixed $\nu = m + 1$" and "Hyperprior") provide the ratio with respect to the benchmark. If a value is greater than 1, it means that the flat prior outperforms the other priors, while if the value is lower than 1, the flat prior shows poorer performance. In Table 1, increasing ν from 5 to 15 leads to strong improvements of around 12% at horizon 3 and 27% at horizon 5 for the proposed loss-based prior against the other priors. When $\nu = 5$, the differences between the hyperprior and the flat prior are small except for horizon 5 and 7, while for $\nu = 10$, the improvement ranges from 7% at horizon 3 to 31% at horizon 7.

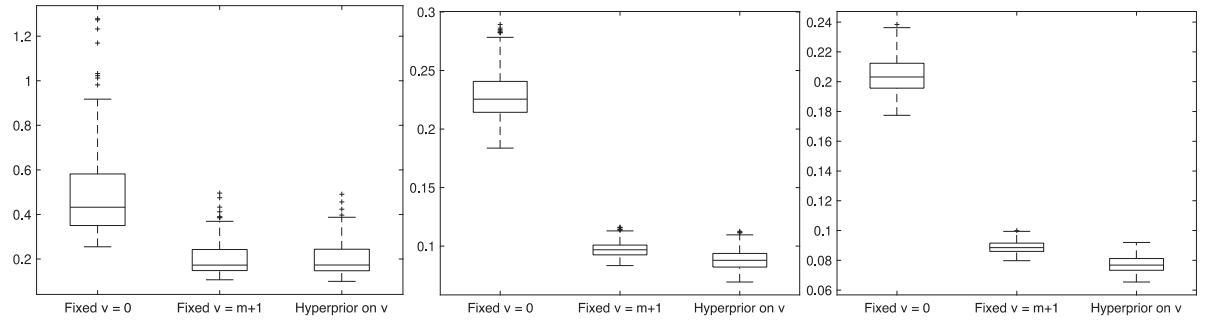

Fig. 4. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 20$ and sample size $T = 30$. Results are reported for data generated from a Wishart distribution with $\nu = 20$ (left), $\nu = 24$ (center), and $\nu = 26$ (right).

Table 2
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 10$ and sample size $T = 30$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 10$			$\nu = 15$			$\nu = 20$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.649	0.996	0.996	0.382	0.999	0.999	0.328	1.00	1.000
3	0.636	1.028	1.029	0.463	0.972	0.951	0.474	0.97	0.930
5	1.247	0.918	0.911	1.51	0.903	0.823	1.841	0.914	0.798
7	3.953	0.863	0.853	5.687	0.855	0.741	7.895	0.863	0.699

Table 3
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 20$ and sample size $T = 30$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 20$			$\nu = 24$			$\nu = 26$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.664	1.003	1.003	0.414	1.002	1.002	0.382	1.001	1.001
3	0.643	1.006	1.007	0.509	0.904	0.895	0.511	0.89	0.878
5	2.447	0.804	0.805	2.686	0.793	0.766	2.839	0.793	0.759
7	12.920	0.734	0.735	15.548	0.718	0.684	17.127	0.719	0.671

These results are also confirmed in high dimensional cases as shown in Figures 3 and 4 for the ten-dimensional and twenty-dimensional cases, respectively. In Figure 3, we compare our loss-based hyperprior with the fixed ν equal to 0 and $m + 1$ for the data generated from a Wishart with degrees of freedom equal to 10 (left panel), 15 (center), and 20 (right). In this scenario, the results demonstrate that our loss-based prior is an improvement over fixed ν for the case of 15 and 20 degrees of freedom, while for ν equal to 10, we have small differences between the three prior representations.

Table 2 provides the RMAD of the IRFs for different horizons for the scenario with $m = 10$. These results confirm conclusions obtained from Figure 3, where our hyperprior outperforms alternatives when degrees of freedom are increased. For ν equal to 10, the loss-based prior shows strong differences when the horizons increase. For the medium and large cases, the loss-based prior outperforms the other two priors with fixed ν of around $5 - 7\%$ at horizon 3 and $25 - 30\%$ at horizon 7.

For the twenty-dimensional case, Figure 4 reports the results for data simulated from a Wishart with 20 (left panel), 24 (center), and 26 (right) degrees of freedom with T equal to 30. Our loss-based prior and the fixed $\nu = m + 1$ prior behave similarly for the left and center panels, with both outperforming the flat prior. When data are generated from a Wishart distribution with 26 degrees of freedom (right panel), our loss-based prior outperforms the fixed $\nu = m + 1$ prior. Table 3 provides the RMAD of the IRFs for the dimensionality $m = 20$ and it confirms the findings from Figure 4. Hence, the improvements are substantial when degrees of freedom and horizons are increased.

These results are confirmed when T is equal to both 100 and 240 as shown in Appendix B.

4. Case study 1: Forecasting macroeconomic data

4.1. Data description

The FRED data are a set of key US macroeconomic quantities sampled at quarterly frequency from the second quarter of 1959 to the third quarter of 2019. All variables are transformed to be stationary using the approach of McCracken and Ng (2020) and three different sets of variables are used to estimate a small, medium, and large-scale VAR model.

The small-scale VAR model considers three variables, which represent inflation, the Gross Domestic Product (GDP), and the target interest rate. In particular, the GDP is measured in Billions of Dollars. Inflation is measured by the GDP deflator, which computes the changes in prices for all goods and services produced in the economy and differs from the Consumer Price Index (CPI) because it is not based on a fixed basket of goods. The final variable is the effective Federal Funds Rate (FEDFUNDS), which is the target interest rate set by the Federal Open Market Committee at which commercial banks borrow and lend their excess reserves to each other overnight.

For the medium-scale model, we additionally consider consumption, investment, and production variables. The real personal consumption expenditure (PCECC96) is a measure of consumer or household spending for a period and it is used to construct the PCE Price index, which measures the price changes in consumer goods and services in the US economy. Real gross private domestic investment (GPDIC1) is a component of the GDP and measures the quantity of money invested by private businesses in the domestic economy. The average weekly hours of production and nonsupervisory employees for the manufacturing sector (AWHMAN) relate to the average hours per worker for which pay was received, and it differs from the standard and scheduled hours.

The large-scale model additionally uses the macroeconomic variables related to GDP, inflation, production, consumption, and investment jointly with private investment in the residential sector (PRFIx), consumption (PCECTPI), and common stock index based on the S&P 500 index (SP500). The Industrial Production Index (INDPRO) measures the level of production and capacity in the manufacturing, mining, electric, and gas industries relative to 2012. Capacity Utilization (CUMFNS) captures the manufacturing and production capabilities that are being used by the economy at any given time, relating the output produced with the given resources and the potential output that can be produced if capacity is fully used. Lastly, CPI for All Urban Consumers (CPIAUCSL) measures the average change over time in the prices paid by consumers for a market basket of consumer goods and services.

4.2. Forecasting results

We evaluate the performance of our loss-based hyperprior with respect to the fixed ν prior by forecasting one-quarter ahead ($h = 1$). We compare the predictive ability of the three different priors by using point and density forecasting measures. To evaluate the forecasting capability, we compute the root mean square error (RMSE), given by

$$RMSE_i = \left[\frac{1}{T-R} \sum_{t=R}^{T-1} (\hat{y}_{i,t+1} - y_{i,t+1})^2 \right]^{\frac{1}{2}}, \tag{11}$$

where R is the length of the rolling window, $y_{i,t+1}$ is the observation for the i-th variable, and $\hat{y}_{i,t+1}$ is the one-step ahead prediction for the i-th variable.

We assess the density forecasting using the continuous ranked probability score (CRPS) introduced by Gneiting and Raftery (2007) and Gneiting and Ranjan (2011). The use of the CRPS has some advantages with respect to the log score since it weights values from the predictive density that are close to the outcome, and it is less sensitive to outlier outcomes. The CRPS is defined such that a lower value indicates better performance, and is given by

$$\begin{aligned} CRPS_t(y_{t+1}) &= \textstyle\int_{-\infty}^{+\infty} (F(z) - \mathbb{1}(y_{t+1} \leq z))^2 dz \\ &= E_f|Y_{t+1} - y_{t+1}| - 0.5 E_f|Y_{t+1} - Y'_{t+1}|, \end{aligned} \tag{12}$$

where $F(\cdot)$ is the cumulative distribution function associated with the posterior predictive density, f, $\mathbb{1}(y_{t+1} \leq z)$ is an indicator function taking the value 1 if $y_{t+1} \leq z$ and 0 otherwise, and Y_{t+1}, Y'_{t+1} are independent random draws from the posterior predictive density.

In addition, we apply Diebold-Mariano t tests (Diebold and Mariano, 1995) for equality of the average loss (with loss defined as the RMSE or CRPS) to compare the predictions of alternative models with the benchmark. The differences in accuracy that are statistically different from zero are denoted with one, two, or three asterisks, corresponding to significance levels of 10%, 5%, and 1%, respectively.

The small-scale VAR only includes three variables, the medium-scale VAR seven variables, and the large-scale VAR 15 variables. Given the quarterly frequency of our data, we include $p = 5$ lags for all the models considered, and we fit the models using the MCMC algorithm with 6000 iterations after discarding the first 1000 iterations as burn-in. For forecasting, we use a rolling window size of 60 quarters.

Figure 5 shows the results of the estimated degrees of freedom jointly with the 95% highest posterior density (HPD) and the degrees of freedom used in the Horseshoe-Wishart scenario with fixed $\nu = m + 1$ (in red).

From Figure 5, we observe strong changes in the estimated degrees of freedom across time. The left panel results indicate an increase in the degrees of freedom after 2000 and a fall around 2009, strongly linked to the Lehman Brothers failure.

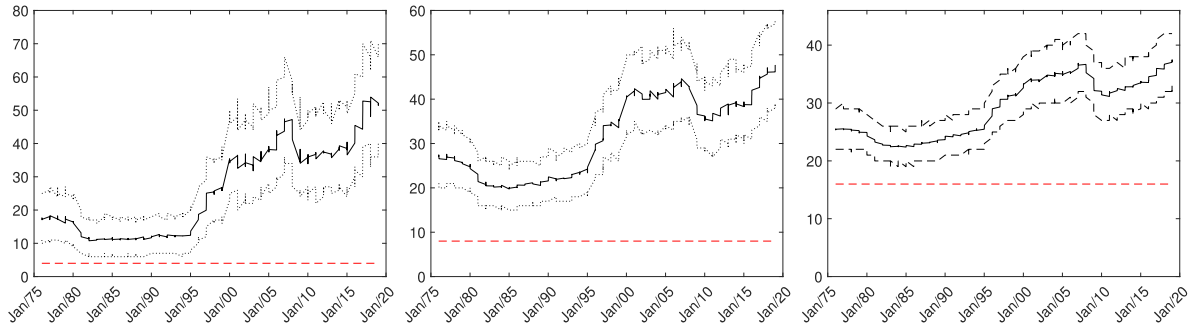

Fig. 5. Estimated degrees of freedom (solid line) for the loss-based hyperprior by using a rolling window of 60 quarters with the 95% Highest Posterior Density (dotted lines) and the case with $\nu = m + 1$, where $m \in \{3, 7, 15\}$ (red dashed line) for the macroeconomic data. The left panel is for the small-case; the center for the medium-scale and the right for the large-scale.

Table 4
RMSE (Columns 2-4) and average CRPS (Columns 5-7) for the small-case VAR for each prior. Column "Fixed $\nu = 0$" provides the RMSE and the average CRPS; Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred prior and the flat prior. ***, ** and * indicate ratios are significantly different from 1 at the 1%, 5% and 10% significance level according to the Diebold-Mariano test.

Variable	RMSE			average CRPS		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
GDPC1	0.0092	0.9459	0.9116***	0.0513	0.9653***	0.8483***
GDPCTPI	0.0044	0.9484	0.9442**	0.0516	0.9616***	0.8423***
FEDFUNDS	1.0206	0.996	1.0012	0.4060	0.9880***	0.9713***

These results are confirmed also in the medium-scale VAR model (center panel) with an increase also observed in the last period. These changes are less evident in the large-scale VAR model (right panel) but provide clear support for not fixing ν. The identified changes in degrees of freedom over time can be associated with the uncertainty (see, Bloom, 2014), illustrating countercyclical fluctuations during recessionary periods.

Table 4 shows the results for the small-case VAR for the RMSE (left panel) and the CRPS (right panel). The column "Fixed $\nu = 0$", which is the benchmark, reports the RMSE and the average CRPS, while the columns "Fixed $\nu = m + 1$" and "Hyperprior", provide the ratio between each prior and the flat prior. When the ratio is less than 1, it indicates that the model with loss-based hyperprior or with fixed $\nu = m + 1$ outperforms the benchmark model with fixed $\nu = 0$. In the point forecasting measure, the benchmark model outperforms our hyperprior for the FEDFUNDS by a small amount. On the other hand, for the GDP, our hyperprior leads to an improvement of around 9% with respect to the benchmark. This result is confirmed for the average CRPS, where our loss-based hyperprior outperforms the benchmark model by 15% for real GDP growth and the GDP deflator, and by 3% for FEDFUNDS. By using the Diebold-Mariano test, we show evidence of statistical significance for all variables when density forecasting is considered, while for point forecasting, the results are statistically significant for GDP and GDP deflator. In addition, our loss-based prior shows improved density forecasting performance relative to the alternative fixed $\nu = m + 1$ prior.

For the 7-variable model, Table 5 presents the point and density measures, and similar results can be observed for both forecasting performance measures. The main difference relates to the GDPCTPI, where the hyperprior approach outperforms the benchmark by about 11% in point forecasting, whilst for the FEDFUNDS, there is little difference. Moreover, we observe that the hyperprior model outperforms the other models for the PCECC96 by about 11%, while for the GDP both the hyperprior and the model with fixed $\nu = m + 1$ (with $m = 7$) demonstrate around 7% and 9% improvement. As in the small-scale VAR, the average CRPS shows better results with respect to the benchmark model across the 7 variables. In particular, for the GDP, GDP deflator, and the FEDFUNDS, the hyperprior model outperforms the benchmark by 4% up to 16%. This is also confirmed for the other variables analysed and from the Diebold-Mariano test for the density forecasting measures.

The results for the large-scale VAR with 15 variables are presented in Table 6. For point forecasting, the three main variables of interest show similar improvements to the one in the medium-scale VAR. In fact, the loss-based hyperprior improves with respect to the benchmark by about 12% for the GDP and 28% for its deflator, while for the FEDFUNDS the situation is similar. The average CRPS demonstrates that the improvement is stronger for every variable, particularly for the FEDFUNDS. If we look at all 15 variables analysed, the loss-based hyperprior model always outperforms the benchmark in a density forecasting scenario, which is also highlighted by the Diebold-Mariano test. These results are confirmed when looking at the prior with fixed ν equal to $m + 1$, where our loss-based prior outperforms the other prior in density forecasting.

Table 5
RMSE (Columns 2-4) and average CRPS (Columns 5-7) for the medium-case VAR for each prior. Column "Fixed $\nu = 0$" provides the RMSE and the average CRPS; Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred prior and the flat prior. ***, ** and * indicate ratios are significantly different from 1 at the 1%, 5% and 10% significance level according to the Diebold-Mariano test.

Variable	RMSE			average CRPS		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
GDPC1	0.0095	0.9081**	0.9284**	0.0823	0.9304***	0.8396***
GDPCTPI	0.0060	0.9155*	0.8888**	0.0834	0.9252***	0.8368***
FEDFUNDS	1.0048	1.0005	1.0002	0.4165	0.9821***	0.9621***
PCECC96	0.0082	0.9796	0.8926**	0.0837	0.9288***	0.8335***
GPDIC1	0.0342	0.9868	1.0029	0.0866	0.9300***	0.8435***
AWHMAN	0.2470	1.0003	0.9995	0.1545	0.9641***	0.9308***
CES2000000008x	0.0086	0.8936	0.9486	0.0859	0.9268***	0.8336***

Table 6
RMSE (Columns 2-4) and average CRPS (Columns 5-7) for the large-case VAR for each prior. Column "Fixed $\nu = 0$" provides the RMSE and the average CRPS; Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred prior and the flat prior. ***, ** and * indicate ratios are significantly different from 1 at the 1%, 5% and 10% significance level according to the Diebold-Mariano test.

Variable	RMSE			average CRPS		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
GDPC1	0.0093	0.9105	0.8833**	0.0905	0.8542***	0.7783***
GDPCTPI	0.0074	0.8672	0.7153***	0.0913	0.8520***	0.7747***
FEDFUNDS	0.9441	1.0057	1.0065	0.4138	0.9613***	0.9454***
PCECC96	0.0085	0.9471	0.9442**	0.0920	0.8502***	0.7781***
GPDIC1	0.0324	1.0003	1.0185	0.0945	0.8529***	0.7827***
AWHMAN	0.2484	0.9978	0.9892*	0.1616	0.9286***	0.8957***
CES2000000008x	0.0091	0.8400***	0.8939**	0.0939	0.8493***	0.7753***
PRFIx	0.0407	0.9914	0.9930	0.0970	0.8520***	0.7813***
INDPRO	0.0146	0.8374***	0.9306***	0.0953	0.8501***	0.7770***
CUMFNS	1.0338	0.9915	0.9844	0.5651	0.9586***	0.9435***
SRVPRD	0.0081	0.9427	0.9696	0.0965	0.8458***	0.7727***
PCECTPI	0.0104	0.7792***	0.6835***	0.0972	0.8484***	0.7703***
GPDICTPI	0.0094	0.9068**	0.8185***	0.0980	0.8435***	0.7703***
CPIAUCSL	0.0111	0.9165	0.8336**	0.0992	0.8452***	0.7668***
SP500	0.0679	0.9913**	1.0188	0.1065	0.8517***	0.7830***

5. Case study 2: Dengue data

The second case study analyses the Google Dengue Trend (GDT) dataset, which tracks the Dengue incidence based on internet search patterns and clusters weekly queries for key terms related to the disease. We use GDT data from January 2011 to December 2014 for Argentina, Bolivia, Brazil, India, Indonesia, Mexico, Philippines, Singapore, Thailand, and Venezuela. Following Davis et al. (2016), we examine a VAR model with two lags, and we run a forecasting exercise with a rolling window of 104 weekly observations.

In Figure 6 we present the posterior mean of the degrees of freedom evaluated using rolling window estimation from the model using our hyperprior (solid black line), the 95% HPD (dotted line), and the fixed values of ν equal to 11. The results indicate that the estimated degrees of freedom are often considerably larger than the fixed value of 11. We observe some changes in values at the beginning of the sample and then a decrease before it remains relatively stationary.

Table 7 shows the forecasting results for each country. For the RMSE, the proposed loss-based hyperprior leads to small improvements with respect to the benchmark prior and the other fixed prior for Bolivia, India, and Mexico, while for Brazil and Philippines, the model with fixed prior equal to $\nu = m + 1$ performs better with respect to the benchmark and slightly better than the proposed loss-based prior. The average CRPS indicates strong improvement against the benchmark model and the fixed ν equal to $m + 1$ prior for all countries. In particular, the proposed loss-based prior outperforms the benchmark by 5% for India, Philippines, and Brazil, and by 3% for Argentina, Indonesia, Thailand, and Venezuela. These results provide evidence of strong significance from density forecasting for all countries as shown by the Diebold-Mariano test. The Diebold-Mariano test gives statistically significant results for only Bolivia, Brazil, and India.

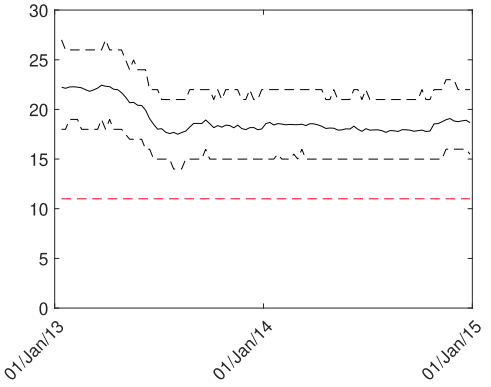

Fig. 6. Estimated degrees of freedom (solid line) for the loss-based hyperprior by using a rolling window of 104 weekly data with the 95% Highest Posterior Density (dotted lines) and the case with $\nu = m + 1$, where $m = 10$ (red dashed line) for the Dengue data.

Table 7
RMSE (Columns 2-4) and average CRPS (Columns 5-7) for the Dengue Data for each prior. Column "Fixed $\nu = 0$" provides the RMSE and the average CRPS; Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred prior and the flat prior. ***, ** and * indicate ratios are significantly different from 1 at the 1%, 5% and 10% significance level according to the Diebold-Mariano test.

Variable	RMSE			average CRPS		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
Argentina	0.3475	1.0033	1.0026	0.1933	0.9864***	0.9761***
Bolivia	0.2529	0.9982	0.9972*	0.1682	0.9711***	0.9601***
Brazil	0.2389	0.9894**	0.9899***	0.1482	0.9618***	0.9529***
India	0.2142	0.9912**	0.9905**	0.1607	0.9673***	0.9494***
Indonesia	0.2696	1.0097	1.0079	0.1651	0.9787***	0.9690***
Mexico	0.2425	0.9997	0.9986	0.1547	0.9749***	0.9593***
Philippines	0.2513	0.9905**	0.9928	0.1532	0.9614***	0.9488***
Singapore	0.7310	1.0068	1.0083	0.3501	0.9923*	0.9912*
Thailand	0.3731	1.0033	1.0028	0.2010	0.9789***	0.9710***
Venezuela	0.2563	1.0055	1.0019	0.1594	0.9831 ***	0.9692***

6. Discussion

We have presented a novel method to perform forecasting in VAR models, where a hyperprior is set on the number of degrees of freedom of the covariance matrix for the global-local-shrinkage-Wishart prior. The proposed loss-based prior takes into consideration only the intrinsic properties of the model; that is, the sampling distribution and the priors. The method has been compared with what is currently used in the literature when no information about the parameter values of the Horseshoe-Wishart prior is available.

The analysis of simulated data has shown that, when the true value of ν is close to $m + 1$, both approaches perform similarly. While, as one would expect, the farther the true ν is from $m + 1$, the better the performance of the proposed prior. To illustrate in practice the advantage of having a loss-based prior on ν, we have analysed its performance in terms of prediction on two datasets. One concerns macroeconomic variables from the FRED dataset, and the other is the analysis of infection variables from the GDT dataset. For both datasets, the proposed method outperforms the one currently used, in particular for density forecasting.

In support of our results, we have estimated the number of degrees of freedom using rolling windows. This analysis has shown that the data contains information for a value of ν always above the value of $m + 1$, which justifies the use of the proposed method. As the data appears to have been generated by a Bayesian model with a relatively large number of degrees of freedom, by setting $\nu = m + 1$, one impacts the predictive performance of the model. On the other hand, by assuming uncertainty of the value of ν by assigning an objective prior to it, the model is free to "choose" the most appropriate value of the parameter and, even considering the extra uncertainty that this implies, the results are better than the previous method.

There are at least two possible research directions to further extend this work. Given the increasing importance of forecasting macroeconomic variables on policymakers' agendas, one may consider incorporating time-varying volatility, such as stochastic volatility, into the VAR framework when estimating the degrees of freedom through a loss-based prior (see, e.g., Clark and Ravazzolo, 2015; Kastner and Huber, 2020). Additionally, another interesting avenue is to explore the application

of the proposed prior in analyzing matrix VAR in the form of the Wishart Autoregressive process of multivariate stochastic volatility (Gourieroux et al., 2009).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Luca Rossini: Methodology, Software, Writing – original draft, Writing – review & editing. **Cristiano Villa:** Methodology, Software, Writing – original draft, Writing – review & editing. **Sotiris Prevenas:** Methodology, Software, Writing – original draft, Writing – review & editing. **Rachel McCrea:** Methodology, Software, Writing – original draft, Writing – review & editing.

Acknowledgments

The authors are grateful to the two referees and the Associate Editor for the constructive comments on earlier versions of the paper which helped in improving the quality of this work. Luca Rossini acknowledges financial support from the Italian Ministry of University and Research (MUR) under the Department of Excellence 2023-2027 grant agreement "Centre of Excellence in Economics and Data Science" (CEEDS).

Appendix A. Loss-based prior

In this section we illustrate the derivation of the loss-based prior in Eq. (2).

Consider a probability distribution $f(x|\theta)$, where $\theta \in \Theta$ is an unknown discrete parameter. The prior mass to be put on θ is derived by considering what is lost if the model $f(x|\theta)$ is removed and it is the true one. The approach associates a *worth* to each parameter value measured by applying a result available in Berk (1966) which states that, if a model is misspecified (i.e. if θ is removed and it is the true value) then the posterior distribution asymptotically accumulates at the value θ' such that the Kullback–Leibler divergence (Kullback and Leibler, 1951) $KL(f(\cdot|\theta)\|f(\cdot|\theta'))$ is minimised. That is, if the true model is removed, the estimation process will asymptotically indicate as the correct model the nearest one, in terms of the Kullback–Leibler divergence, i.e., the model which is the most similar to the true one (Bernardo and Smith, 1994). To link the *worth* of each parameter value to the prior probability, we use the *self-information* loss function. This particular type of loss function assigns a loss to a probability statement and, say we have defined prior $\pi(\theta)$, its form is $-\log\pi(\theta)$. More information about the self-information loss function can be found, for example, in Merhav and Feder (1998).

To formally derive the prior for θ, we can proceed in terms of utilities, instead of losses; this approach allows for a clearer exposition and does not impact the logic behind the prior derivation. Let us then write utility $u_1(\theta) = \log\pi(\theta)$. We then let the minimum divergence from θ be represented by utility $u_2(\theta)$. We naturally want $u_1(\theta)$ and $u_2(\theta)$ to be matching utility functions; though as it stands $-\infty < u_1 \leq 0$ and $0 \leq u_2 < \infty$, and we want $u_1 = -\infty$ when $u_2 = 0$. The scales are matched by taking exponential transformations, so $\exp(u_1)$ and $\exp(u_2) - 1$ are on the same scale and we obtain

$$\pi(\theta) = e^{u_1(\theta)} \propto e^{u_2(\theta)} - 1,$$

yielding the loss-based prior for θ in Eq. (2).

Appendix B. Further simulation results

B1. Case $T = 100$

In this section, we provide further simulation results. We report the root mean absolute deviation (RMAD) for the case with $T = 100$. In particular, Figure B.7 shows the RMAD for the covariance matrix for the five-dimensional case, when the data are simulated from a Wishart distribution with degrees of freedom equal to 5 (left), 10 (center), and 15 (right). Table B.8 provides the RMAD for the impulse response function at four different horizons $h = 1, 3, 5$ and 7.

In Figure B.8, we show results for a ten-dimensional case for the matrix of covariance and with data generated with 10 (left), 15 (center), and 20 (right) degrees of freedom. Table B.9 shows the IRF for the same dataset for four different horizons.

In conclusion, Figure B.9 shows the results for the twenty-dimensional case, where the data are generated from a Wishart with 20 (left), 24 (center), and 26 (right) degrees of freedom. Table B.10 shows the RMAD for the impulse response functions at four horizons.

As stated in the paper, the results show improvements in the use of our loss-based prior with respect to a fixed ν prior when the data are generated with a higher than the dimension degrees of freedom.

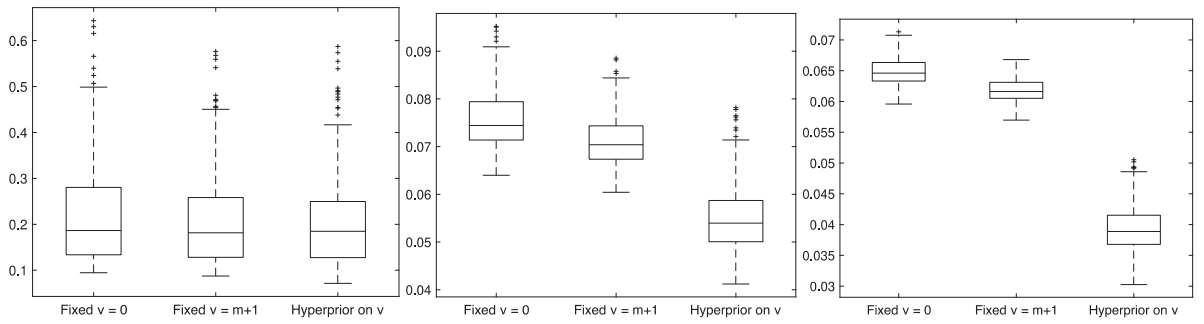

Fig. B7. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 5$ and sample size $T = 100$. Results are reported for data generated from a Wishart distribution with $\nu = 5$ (left), $\nu = 10$ (center), and $\nu = 15$ (right).

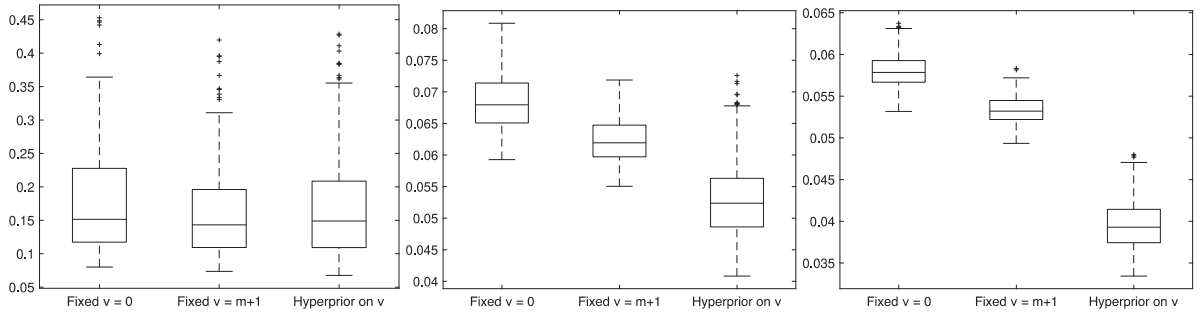

Fig. B8. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 10$ and sample size $T = 100$. Results are reported for data generated from a Wishart distribution with $\nu = 10$ (left), $\nu = 15$ (center), and $\nu = 20$ (right).

Table B8
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 5$ and sample size $T = 100$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 5$			$\nu = 10$			$\nu = 15$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.411	0.996	0.995	0.283	0.998	0.989	0.272	0.999	0.990
3	0.486	1.001	1.001	0.415	0.988	0.982	0.316	0.994	0.933
5	0.424	0.996	0.994	0.331	0.986	0.918	0.438	0.980	0.805
7	0.303	0.997	0.982	0.440	0.972	0.841	0.739	0.967	0.694

Table B9
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 10$ and sample size $T = 100$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 10$			$\nu = 15$			$\nu = 20$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.540	0.996	0.993	0.346	0.997	0.990	0.309	0.999	0.992
3	0.588	1.006	1.007	0.342	0.993	0.976	0.342	0.989	0.945
5	0.631	0.989	0.982	0.524	0.959	0.857	0.686	0.958	0.786
7	0.996	0.960	0.936	1.096	0.933	0.768	1.711	0.938	0.689

B2. Case $T = 240$

As a third simulated experiment, we report the RMAD for the case with $T = 240$. Figure B.10 shows the RMAD for the covariance matrix for the three-dimensional case when the data are generated from a Wishart distribution with degrees of freedom equal to 3 (left), 5 (center), and 7 (right). Table B.11 shows the RMAD for the impulse response variable at four different horizons $h = 1, 3, 5$, and 7.

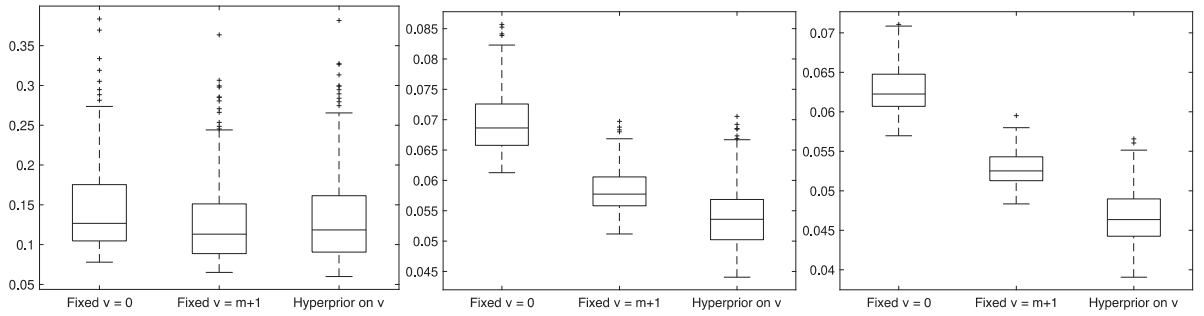

Fig. B9. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 20$ and sample size $T = 100$. Results are reported for data generated from a Wishart distribution with $\nu = 20$ (left), $\nu = 24$ (center), and $\nu = 26$ (right).

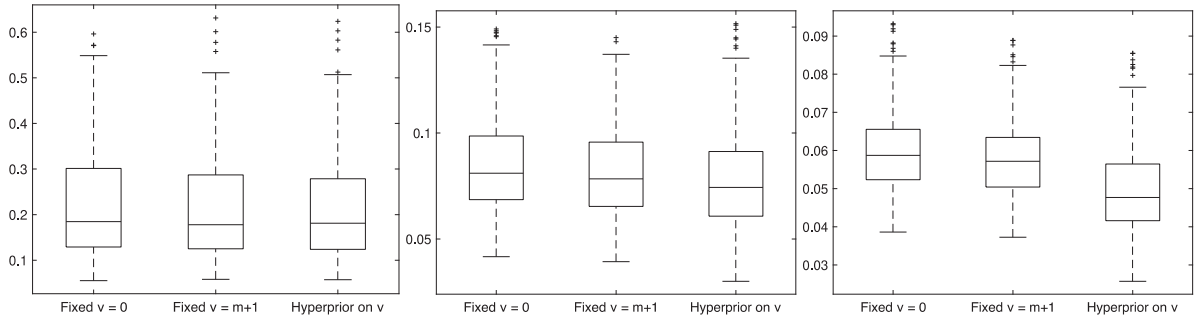

Fig. B10. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 3$ and sample size $T = 240$. Results are reported for data generated from a Wishart distribution with $\nu = 3$ (left), $\nu = 5$ (center), and $\nu = 7$ (right).

Table B10
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 20$ and sample size $T = 100$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 20$			$\nu = 24$			$\nu = 26$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.620	0.996	0.995	0.401	0.997	0.995	0.368	0.998	0.995
3	0.644	1.021	1.025	0.391	0.989	0.982	0.375	0.986	0.971
5	1.193	0.960	0.951	1.061	0.928	0.868	1.137	0.932	0.858
7	3.809	0.923	0.901	3.789	0.893	0.807	4.247	0.899	0.791

Table B11
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 3$ and sample size $T = 240$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 3$			$\nu = 5$			$\nu = 7$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.215	0.999	0.999	0.175	0.999	0.997	0.175	0.999	0.996
3	0.299	1.000	1.001	0.228	1.000	1.000	0.208	1.000	0.997
5	0.240	1.000	1.001	0.193	0.999	0.998	0.185	0.999	0.992
7	0.202	0.999	1.000	0.174	0.998	0.995	0.171	0.998	0.982

In Figure B.11, we study the seven-dimensional case for the matrix of covariances and data generated with 7 (left), 10 (center), and 13 (right) degrees of freedom. The RMAD of the impulse response functions for four different horizons are reported in Table B.12.

In conclusion, Figure B.12 shows the results for the fifteen-dimensional case, where the data are generated from a Wishart with 15 (left); 20 (center), and 25 (right) degrees of freedom. Table B.13 shows the RMAD for the impulse response function for four different horizons.

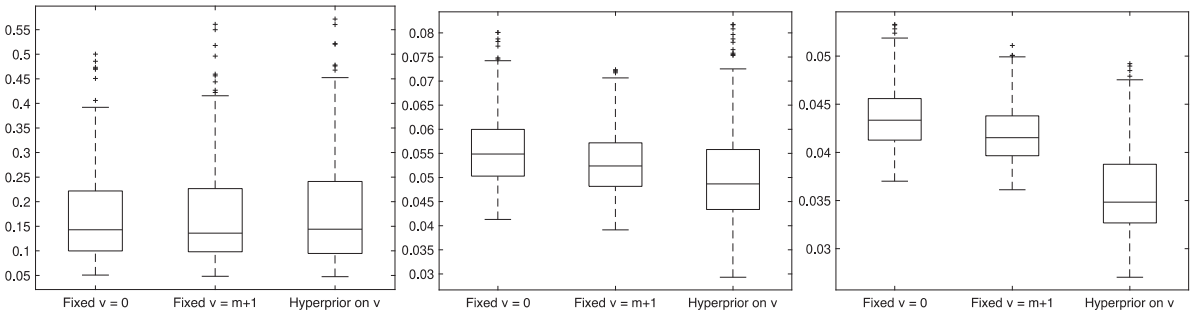

Fig. B11. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 7$ and sample size $T = 240$. Results are reported for data generated from a Wishart distribution with $\nu = 7$ (left), $\nu = 10$ (center), and $\nu = 13$ (right).

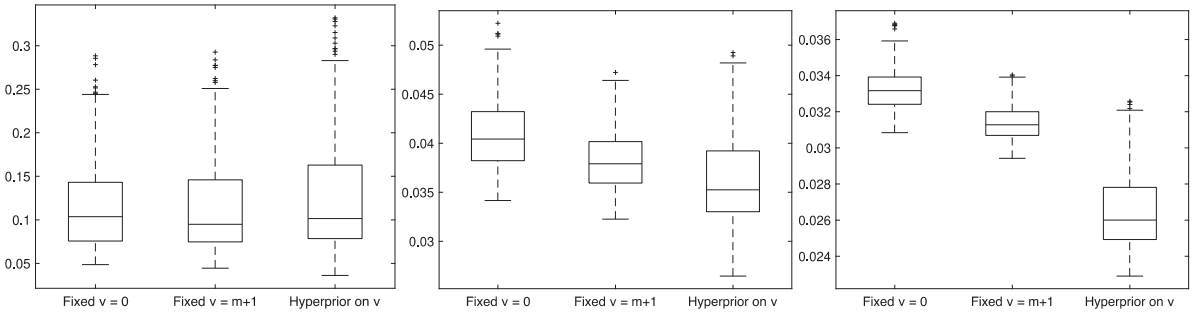

Fig. B12. Monte Carlo Simulation - RMAD of the covariance matrices. These distributions are obtained by simulating 250 VAR(1) with dimension $m = 15$ and sample size $T = 240$. Results are reported for data generated from a Wishart distribution with $\nu = 15$ (left), $\nu = 20$ (center), and $\nu = 25$ (right).

Table B12
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 7$ and sample size $T = 240$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 7$			$\nu = 10$			$\nu = 13$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.474	0.998	0.996	0.273	0.998	0.990	0.262	0.9990	0.988
3	0.582	1.001	1.000	0.275	0.999	0.997	0.252	0.999	0.991
5	0.516	0.999	0.997	0.245	0.997	0.987	0.238	0.995	0.966
7	0.521	0.995	0.995	0.242	0.993	0.966	0.257	0.989	0.924

Table B13
Monte Carlo Simulation - RMAD of the IRF for four horizons $h = 1, 3, 5$ and 7 by simulating 250 VAR(1) with dimension $m = 15$ and sample size $T = 240$. Column "Fixed $\nu = 0$" provides the RMAD of the IRF, while Columns "Fixed $\nu = m + 1$" and "Hyperprior" provide the ratio between the referred priors and the flat prior.

Horizon	$\nu = 15$			$\nu = 20$			$\nu = 25$		
	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior	Fixed $\nu = 0$	Fixed $\nu = m + 1$	Hyperprior
1	0.567	0.997	0.994	0.337	0.997	0.988	0.302	0.998	0.988
3	0.666	1.004	1.007	0.310	0.999	0.993	0.285	0.995	0.973
5	0.691	0.994	0.993	0.354	0.981	0.917	0.418	0.972	0.835
7	1.069	0.975	0.960	0.588	0.958	0.820	0.855	0.952	0.727

Appendix C. Convergence Diagnostics

In this section of the appendix, we describe the converge analysis we have performed for the different simulation experiments.

The convergence analysis has been done using the R Coda Package (Plummer et al., 2006). In particular, we provide the chain of the degrees of freedom, the Geweke convergence test, the Gelman-Rubin's plot, and test statistics. Figure C.13 provides the chain of the degrees of freedom when $T = 30$ and m is equal to 5 (left), 10 (center), and 20 (right). As shown

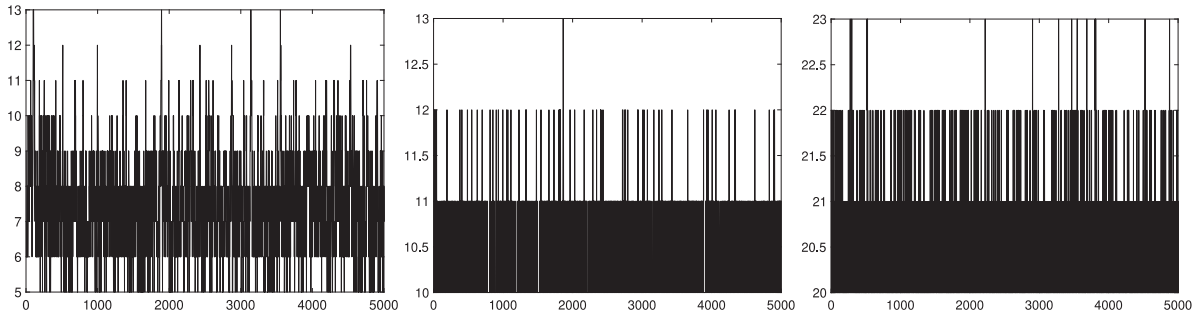

Fig. C13. Posterior Chain of the estimated degrees of freedom when $T = 30$ and m is equal to 5 (left); 10 (center), and 20 (right).

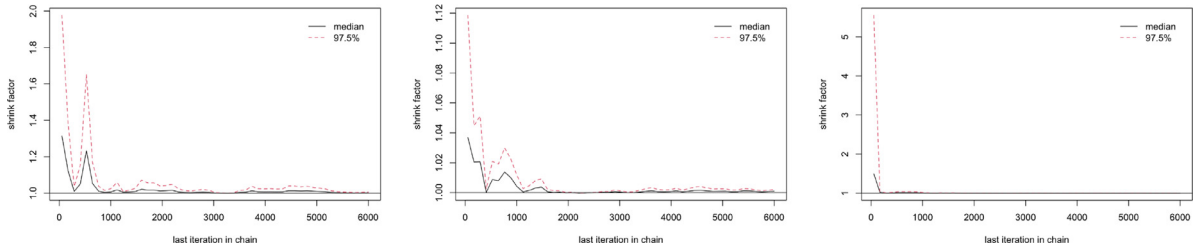

Fig. C14. GelmanRubin plot for the estimated degrees of freedom when $T = 30$ and m is equal to 5 (left); 10 (center), and 20 (right) when no burn-in iterations are discarded.

Table C13
Geweke's test statistics for the posterior chain of the degrees of freedom for different simulation experiments when $T = 30$.

Case: Simulated data	Test
$m = 5$	0.6713
$m = 10$	0.9306
$m = 20$	-0.8532

in the Figure, we are able to reach convergence of the chain over the number of iterations. Table C.13 reports the results of Geweke's convergence test (Geweke, 1992) for the different simulation experiments. The test statistics show no convergence issues. To perform the Gelman-Rubin test of convergence, we have run multiple chains with sparse starting points. Figure C.14 plots the shrinking factor for different simulation experiments when the burn-in iterations are not discarded and in all cases, we do not see any indication of failed convergence under the proposed loss-based prior.

References

Abramowitz, M., Stegun, I.A., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York.
Bańbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. Journal of Applied Econometrics 25 (1), 71–92.
Berk, R., 1966. Limiting behaviour of posterior distributions when the model is incorrect. Annals of Mathematical Statistics (37) 51–58.
Bernardi, M., Bianchi, D., Bianco, N., 2024. Variational Inference for Large Bayesian Vector Autoregressions. Journal of Business & Economic Statistics 1–17.
Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. Wiley, New York.
Bloom, N., 2014. Fluctuations in Uncertainty. Journal of Economic Perspectives 28 (2), 153–176.
Carneiro, H.A., Mylonakis, E., 2009. Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks. Clinical Infectious Diseases 49 (10), 1557–1564.
Carriero, A., Chan, J., Clark, T.E., Marcellino, M., 2022. Corrigendum to "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors"[J. Econometrics 212 (1) (2019) 137–154]. Journal of Econometrics 227 (2), 506–512.
Carriero, A., Clark, T.E., Marcellino, M., 2015. Bayesian VARs: Specification Choices and Forecast Accuracy. Journal of Applied Econometrics 30 (1), 46–73.
Carriero, A., Clark, T.E., Marcellino, M., 2019. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. Journal of Econometrics 212 (1), 137–154.
Carvalho, C.M., Polson, N.G., Scott, J.G., 2010. The horseshoe estimator for sparse signals. Biometrika 97 (2), 465–480.
Clark, T.E., Ravazzolo, F., 2015. Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. Journal of Applied Econometrics 30 (4), 551–575.
Cross, J.L., Hou, C., Poon, A., 2020. Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. International Journal of Forecasting 36 (3), 899–915.
Davis, R.A., Zang, P., Zheng, T., 2016. Sparse Vector Autoregressive Modeling. Journal of Computational and Graphical Statistics 25 (4), 1077–1096.
De Mol, C., Giannone, D., Reichlin, L., 2008. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? Journal of Econometrics 146 (2), 318–328.
Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–263.
Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Reviews 3 (1), 1–100.
Elezovic, N., Giordano, C., Pecaric, J., 2000. The best bounds in Gautschi's inequality. Mathematical Inequalities & Applications 3 (2), 239–252.
Follett, L., Yu, C., 2019. Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior. Econometrics and Statistics 11, 130–144.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J.M., Berger, J., Dawid, A.P., Smith, J.F.M. (Eds.), Bayesian Statistics 4. Oxford University Press, Oxford, pp. 169–193.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association 102 (477), 359–378.
Gneiting, T., Ranjan, R., 2011. Comparing density forecasts using threshold-and quantile-weighted scoring rules. Journal of Business & Economic Statistics 29 (3), 411–422.
Gourieroux, C., Jasiak, J., Sufana, R., 2009. The Wishart Autoregressive process of multivariate stochastic volatility. Journal of Econometrics 150 (2), 167–181.
Gruber, L., Kastner, G., 2023. Forecasting macroeconomic data with Bayesian VARs: Sparse or dense? It depends!. arXiv preprint arXiv:2206.04902.
Huber, F., Feldkircher, M., 2019. Adaptive Shrinkage in Bayesian Vector Autoregressive Models. Journal of Business & Economic Statistics 37 (1), 27–39.
Kastner, G., Huber, F., 2020. Sparse Bayesian vector autoregressions in huge dimensions. Journal of Forecasting 39 (7), 1142–1165.
Koop, G., Korobilis, D., 2010. Bayesian multivariate time series methods for empirical macroeconomics. Foundations and Trends in Econometrics 3 (4), 267–358.
Koop, G., Korobilis, D., 2013. Large time-varying parameter VARs. Journal of Econometrics 177 (2), 185–198.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Annals of Mathematical Statistics (22) 79–86.
Litterman, R.B., 1986. Forecasting with Bayesian vector autoregressions – five years of experience. Journal of Business & Economic Statistics 4 (1), 25–38.
Makalic, E., Schmidt, D.F., 2016. A Simple Sampler for the Horseshoe Estimator. IEEE Signal Processing Letters 23 (1), 179–182.
McCracken, M.W., Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. Journal of Business & Economic Statistics 34 (4), 574–589.
McCracken, M.W., Ng, S., 2020. FRED-QD: A Quarterly database for macroeconomic research. Technical Report. Federal Reserve Bank of St. Louis.
Merhav, N., Feder, M., 1998. Universal prediction. IEEE Transactions on Information Theory 44 (6), 2124–2147.
Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6 (1), 7–11.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.
Sims, C.A., Zha, T., 1998. Bayesian methods for dynamic multivariate models. International Economic Review 39 (4), 949–968.
Stock, J.H., Watson, M.W., 2002. Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business & Economic Statistics 20 (2), 147–162.
Strauss, R.A., Castro, J.S., Reintjes, R., Torres, J.R., 2017. Google dengue trends: An indicator of epidemic behavior. The Venezuelan Case. International Journal of Medical Informatics 104, 26–30.
Uhlig, H., 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics 52 (2), 381–419.
Villa, C., Walker, S.G., 2015. An objective approach to prior mass functions for discrete parameter spaces. Journal of the American Statistical Association 110 (511), 1072–1082.