Macroeconomic forecasting in a multi-country context

Published Version:
Bai, Y., Carriero, A., Clark, T.E., Marcellino, M. (2022). Macroeconomic forecasting in a multi-country context. JOURNAL OF APPLIED ECONOMETRICS, 37(6), 1230-1255 [10.1002/jae.2923].

Availability:
This version is available at: https://hdl.handle.net/11585/900989 since: 2022-11-09

Published:
DOI: http://doi.org/10.1002/jae.2923

Macroeconomic Forecasting in a Multi-country Context[⋆,⋆⋆]

Yu Bai[a], Andrea Carriero[b], Todd E. Clark[c], Massimiliano Marcellino[d]

[a]*Bocconi University*
[b]*Queen Mary University of London and University of Bologna*
[c]*Federal Reserve Bank of Cleveland*
[d]*Bocconi University, IGIER, and CEPR*

Abstract

In this paper we propose a hierarchical shrinkage approach for multi-country VAR models. In implementation, we consider three different scale mixtures of Normals priors — specifically, Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma priors. We provide new theoretical results for the Normal-Gamma prior. Empirically, we use a quarterly data set for the G7 economies to examine how model specifications and prior choices affect the forecasting performance for GDP growth, inflation, and a short-term interest rate. We find that hierarchical shrinkage, particularly as implemented with the Horseshoe prior, is very useful in forecasting inflation. It also has the best density forecast performance for output growth and the interest rate. Adding foreign information yields benefits, as multi-country models generally improve on the forecast accuracy of single-country models.

Keywords: Multi-country VARs; Macroeconomic forecasting; Hierarchical shrinkage; Scale mixtures of Normals priors

JEL classification: C11; C33; C53; C55

⋆This version: January 2022

⋆⋆We would like to thank the editor Herman van Dijk, three anonymous referees, Florian Huber, Gary Koop and participants at the USC Panel Data Forecasting Conference, 2019 Asian Meeting of the Econometrics Society, 25th International Panel Data Conference, 2020 Econometric Society World Congress and seminar participants at Strathclyde Business School for their useful comments. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Bai and Marcellino thank MIUR–PRIN Bando 2017 – prot. 2017TA7TYC for financial support for this research.

1. Introduction

Since the seminal studies by Doan, Litterman, and Sims (1984) and Litterman (1986), Bayesian vector autoregressions (VARs) have become workhorse models in macroeconomic forecasting. Reduced-form VARs are richly parameterized, which brings the risk of overfitting the data and large uncertainty for the future path projected by the model. It is well known that shrinkage generally improves forecasting performance, and Bayesian methods offer an effective way to shrink parameters by using prior information.

Due to increasing international trade and financial flows in recent decades, individual countries are more and more interlinked, which may make it helpful to use multi-country forecasting models and methods. The VAR literature includes three main approaches: (1) factor-augmented VAR models; (2) global vector autoregressive (GVAR) models; and (3) multi-country VARs. In factor-augmented VAR models, each country-specific VAR is augmented with "foreign variables," constructed by using principal components to extract common factors from all variables in foreign countries. In GVARs, used in studies such as Pesaran, Schuermann, and Smith (2009), Cuaresma, Feldkircher, and Huber (2016), Huber (2016), and Dovern, Feldkircher, and Huber (2016), each country-specific VAR is augmented with weakly exogenous "foreign variables," constructed by aggregating other countries' variables with international trade flows as weights. Then, country-specific models are combined to form a global model for the forecasting exercise. In multi-country VARs, used in studies such as Canova, Ciccarelli, and Ortega (2007), Giannone and Reichlin (2009), Korobilis (2016), Dées and Güntner (2017), and Koop and Korobilis (2019), variables for multiple countries are jointly modeled, with various degrees of cross-country interactions. Shrinkage is imposed to deal with the curse of dimensionality and is performed either by considering the panel dimension in the data or by simply treating the multi-country model as a large-scale BVAR and specifying priors on model coefficients.

While conventional Minnesota-type priors are shown to be useful in macroeconomic forecasting and are still widely used in the literature, other work suggests instead applying scale mixtures of Normals priors or other alternatives on single-country BVARs. These prior specifications have advantages with respect to the Minnesota-type prior, since they involve less hyperparameter tuning. They are also computationally more efficient than spike-and-slab priors while enjoying similarly nice theoretical properties at the same time. Huber and Feldkircher (2019) propose applying the Normal-Gamma prior, originally introduced by Griffin and Brown (2010), to BVARs and show that it is beneficial for macroeconomic forecasting. Follett and Yu (2019) use the Horseshoe prior, popular in the statistical literature (Carvalho, Polson, and Scott (2010)), and find that it improves forecast accuracy in a single-country context. Cadonna, Frühwirth-Schnatter, and Knaus (2020) propose a more general Normal-Gamma-Gamma prior, originated in Griffin and Brown (2017), for time-varying parameter BVARs and find that it delivers more sparse parameter estimates (but their study does not address forecast performance). Recently, there has been some limited interest in applying these types of priors in multi-country VARs. Korobilis (2016) proposes the stochastic search specification selection prior, which is a modification of the stochastic search variable selection prior proposed by George, Sun, and Ni (2008) applicable under some model restrictions. Korobilis finds that, to design priors for better forecasting performance, it is important to consider the panel structure in the data. However, the prior specifications considered there introduce dependence across equations, which makes efficient estimation, such as the approach of Carriero, Clark, and Marcellino (2019), difficult to apply.

In this paper, we propose and examine the use of hierarchical shrinkage approaches in multi-country VARs used for macroeconomic forecasting. In implementation, we use three different scale mixtures of Normals priors that have been shown to be successful in single-country BVARs but have not been examined in multi-country models. These priors include the Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma specifications. The hierarchical shrinkage is able to handle the restrictions suggested in Canova and Ciccarelli (2013) for multi-country VARs. It is shown to be computationally more efficient than the existing stochastic search specification selection prior and also delivers better forecasting performance than the existing alternatives. We also provide some novel theoretical results for the Normal-Gamma prior.

Empirically, we work with a quarterly Group of Seven (G7) data set to examine the (point and density) forecasting ability of the new priors for three key macroeconomic variables: output growth, inflation, and a short-term interest rate. We also compare forecasting accuracy across various models with other specification choices. These models include: (1) country-specific VARs, either with Minnesota-type priors or hierarchical shrinkage proposed by Chan (2021); (2) country-specific factor-augmented VARs; (3) GVARs; and (4) multi-country VARs in which shrinkage is performed either by imposing a particular hierarchical factor structure on the model parameters or by using priors. Because stochastic volatility (SV) has been found to be widely useful in macroeconomic forecasting with single-country models and also improves performance in our results, we include SV in all of our model specifications. In addition, since Cross, Hou, and Poon (2020) have raised some questions on the usefulness of scale mixtures of Normals priors in single-country macroeconomic forecasting, we consider alternative hierarchical shrinkage approaches for a robustness check (the appendix includes these results in the multi-country context).

Our results show that hierarchical shrinkage of multi-country VARs, particularly as implemented with the Horseshoe prior, improves macroeconomic forecast accuracy. It has outright advantages for inflation forecasting, and the Horseshoe specification of a multi-country VAR also performs best in density forecasts of output growth and the interest rate.[1] In point forecast accuracy, the Normal-Gamma prior performs best for output growth, whereas the factor shrinkage approach of Canova and Ciccarelli (2009) for multi-country models performs best for the interest rate. These results indicate that, although the Normal-Gamma-Gamma prior is more flexible than the Horseshoe prior and serves as a heavy-tailed extension of the Normal-Gamma prior, these advantages do not yield consistently better forecasting performance. We also find that modeling cross-country interactions achieves gains, as multi-country models generally

[1] Koop and Korobilis (2019) and Feldkircher, et al. (2021) also find that multi-country VARs are more beneficial for inflation forecasting.

outperform single-country models. As is common in the literature, we also find that models' forecasting performance varies over both countries and time. There are countries in which alternative models and priors do better than our hierarchical shrinkage of a multi-country VAR implemented with the Horseshoe prior, but there are no consistent patterns in which one of the alternatives is clearly better. In the interest of brevity, the paper's appendix provides results confirming that stochastic volatility is one important feature to improve forecast accuracy in the multi-country context.

The paper is structured as follows. Section 2 briefly introduces multi-country VAR models, existing prior specifications, and their challenges. Section 3 provides our new hierarchical shrinkage approach for multi-country VARs, the prior specifications, and some new theoretical results. Section 4 gives a brief summary of estimation algorithms and highlights some computational comparisons. Section 5 describes the data, forecasting metrics, and design of our forecasting exercise. Section 6 presents the main empirical results. Section 7 concludes. Technical details, sampling algorithms for various models, and additional empirical results are provided in the appendix.

2. Multi-country VARs

2.1. The model

The multi-country VAR model we consider has the form

$$y_{i,t} = c_i + B_i(L)Y_{t-1} + u_{i,t}, \tag{1}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$; $y_{i,t}$ is a $G \times 1$ vector of variables for each country i, and $Y_t = (y'_{1,t}, \ldots, y'_{N,t})'$; c_i is a $G \times 1$ vector of constant terms for each i, $B_i(L) = \sum_{\ell=1}^{p} B_{i,\ell} L^{\ell}$, where $B_{i,\ell}$ are $G \times NG$ coefficient matrices associated with lag ℓ, $\ell = 1, ..., p$; and $u_{i,t}$ is a $G \times 1$ vector of disturbances. The lag length is assumed to be p. Combining equations across countries, the VAR can be written in matrix form as:

$$Y_t = c + \sum_{\ell=1}^{p} B_\ell Y_{t-\ell} + u_t, \tag{2}$$

where $c = (c'_1, c'_2, \ldots, c'_N)'$, each B_ℓ has dimension $NG \times NG$, and $u_t = (u'_{1,t}, \ldots, u'_{N,t})'$. For the stochastic volatility (SV) specification, we assume that

$$u_t = A^{-1} H_t^{0.5} \epsilon_t, \epsilon_{i,t} \sim i.i.d.\ N(0, I_{NG}),$$

where A^{-1} is a lower triangular matrix with diagonal elements equal to 1, and H_t is diagonal with generic j-th element $h_{j,t}$ evolving as a random walk (RW):[2]

$$\ln h_{j,t} = \ln h_{j,t-1} + e_{j,t},\ j = 1, \ldots, NG, \tag{3}$$

where $e_t = (e_{1,t}, e_{2,t}, \ldots, e_{NG,t})'$ and $e_t \sim N(0, \Phi)$ with a full covariance matrix Φ as in Primiceri (2005). The reduced-form error covariance matrix is $\Sigma_t = A^{-1}H_tA^{-1\prime}$.

While some work has examined time-varying coefficient VAR models (see, e.g., Cogley and Sargent (2005); Primiceri (2005); Koop, Leon-Gonzalez, and Strachan (2009); and D'Agostino, Gambetti, and Giannone (2013)), we restrict attention to constant coefficient VAR models with stochastic volatility for two reasons. First, time-varying coefficient VAR models are rarely used with more than 4-5 variables. This is mainly due to computational complexity and makes recursive forecasting with MCMC methods computationally infeasible. Second, in a forecasting context, reaching parsimony (in terms of both controlling time variation and getting rid of irrelevant regressors) in large models with time-varying coefficients remains a challenging task.[3]

2.2. *Existing priors*

The specification in (2) can incorporate complex dynamic structures for each variable in different countries. However, it also suffers from the curse of dimensionality due to the high dimensionality of the parameter space. For instance, in the forecasting exercises we use data on 3 dependent variables ($G = 3$) for the G7 countries ($N = 7$) and four lags ($p = 4$). A multi-country VAR with such choices would have 1,785 VAR coefficients. Thus, shrinkage is desirable.

In a single-country framework, the macro VAR literature generally relies on Bayesian shrinkage by imposing a Minnesota-type prior (Litterman (1986)) on the VAR coefficients. Applied analogously in our multi-country setup, the prior for B is $\text{vec}(B) \sim N(\text{vec}(\underline{\mu}_B), \underline{\Omega}_B)$, and $\underline{\Omega}_B$ is set to

$$\text{Var}(B_\ell^{(ii)}) = \frac{\lambda_1}{\ell^{\lambda_3}}, \quad \ell = 1, \ldots, p \tag{4}$$

$$\text{Var}(B_\ell^{(ij)}) = \frac{\lambda_2}{\ell^{\lambda_3}} \frac{\sigma_i^2}{\sigma_j^2}, \quad \forall i \neq j, \ell = 1, \ldots, p, \tag{5}$$

[2] The RW specification may come at the cost of generating excessively thick forecast densities. Alternatively, the SV process (3) could be specified as an $AR(1)$ process, and $e_{ij,t}$ could be assumed to be t-distributed to incorporate fat tails. However, Clark and Ravazzolo (2015) find that these alternative specifications fail to dominate a baseline RW specification.

[3] To have parameter time variation in large VARs, Koop and Korobilis (2019) introduce forgetting factors, and Kapetanios, Marcellino, and Venditti (2019) use non-parametric methods combined with stochastic coefficient constraints. Yet, both methods become computationally infeasible if combined with the commonly used stochastic volatility specification. Gefang, Koop, and Poon (2022) develop variational Bayes methods (which utilize approximations of conventional posteriors) that permit large models with time-varying parameters and volatilities. Empirically, since forecasts are computed recursively, we implicitly consider the potential time variation of parameters in the model.

where $B_\ell^{(ij)}$ denotes the element in row i and column j of the matrix B_ℓ, $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$ is the collection of prior hyperparameters, and σ_i^2, σ_j^2 are local scale parameters. For each element i of the intercept vector c, it is common to specify an uninformative prior by setting the prior variance equal to $100 \times \sigma_i^2$.

In view of the fact that the usual Minnesota-type prior ignores the panel structure in the data, Angelini, et al. (2019) recently proposed a modified Minnesota-type shrinkage prior to carefully deal with the panel structure. In particular, a different hyperparameter λ_4 is introduced in (5) on coefficients related to other countries' variables. Angelini, et al. (2019) apply this approach in a forecasting exercise with a Euro area data set and find that it provides some gains. However, this still belongs to the class of Minnesota-type priors. This may come with costs due to parameter uncertainty, since the hyperparameters $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and local scale parameters σ_i^2, σ_j^2 (commonly obtained from AR(1) estimates) are set to some pre-specified values in the estimation.[4]

Since there are a variety of restrictions of interest in multi-country VARs, another strand of the literature suggests designing shrinkage priors to explore these restrictions. Consider the coefficient matrix B_l defined in (2):

$$B_\ell = \begin{bmatrix} B_{11,\ell} & \cdots & B_{1N,\ell} \\ \vdots & \ddots & \vdots \\ B_{N1,\ell} & \cdots & B_{NN,\ell} \end{bmatrix}, \ell = 1, \dots, p, \tag{6}$$

where each block $B_{ij,\ell}$, $i, j = 1, \dots, N$, has dimension $G \times G$. According to Canova and Ciccarelli (2013), it is interesting to check whether certain restrictions exist and what their implications are. For example, cross-sectional heterogeneity (CSH) exists when $\exists\ i, j, i \neq j$, such that $B_{ii,l} \neq B_{jj,l}$ for some l and $c_i \neq c_j$. Dynamic interdependencies (DI) occur when at least one block $B_{ij,l} \neq 0$ for a given i, l and $i \neq j$.[5]

In a special case with one lag ($p = 1$) and no SV, Koop and Korobilis (2016) develop the stochastic model specification search (SSSS) prior:

$$vec(B_{ij,1}^{\text{DI}}) \sim (1 - \gamma_{ij}^{\text{DI}}) N(0, \tau_{ij}^2 \times \underline{c}^{\text{DI}} \times I_G) + \gamma_{ij}^{\text{DI}} N(0, \tau_{ij}^2 \times I_G), \quad i, j = 1, \dots, N, i \neq j \tag{7}$$

$$vec(B_{ii,1}^{\text{CSH}}) \sim (1 - \gamma_{ii}^{\text{CSH}}) N(vec(B_{jj,1}), \xi_{ij}^2 \times \underline{c}^{\text{CSH}} \times I_G) + \gamma_{ii}^{\text{CSH}} N(vec(B_{jj,1}), \xi_{ij}^2 \times I_G), \quad i, j = 1, \dots, N, \tag{8}$$

where $\gamma^{\text{DI}}, \gamma^{\text{CSH}}$ are indicators, τ_{ij}^2, ξ_{ij}^2 are prior variance parameters, and $\underline{c}^{\text{DI}}, \underline{c}^{\text{CSH}}$ are small constants to make prior variances smaller in the spike components. The priors in (7)-(8) provide an extension of the stochastic search variable selection (SSVS) prior in George, Sun, and Ni (2008) to multi-country VARs. It is the first attempt to examine the existence (or absence) of certain dependencies and homogeneities for

[4] Angelini, et al. (2019) also consider a conjugate Minnesota prior in which they set hyperparameters by maximizing the marginal likelihood, as in Giannone, Lenza, and Primiceri (2015). For the non-conjugate prior, they set $\lambda_2 = 1$, $\lambda_4 = 1/2$, and the others to the posterior mode of the values obtained by the methodology of Giannone, Lenza, and Primiceri (2015).

[5] There is one other class of possibly important restrictions: static interdependencies (SI), which occur when the covariance matrix Σ_t is not block diagonal. However, it is not easy to implement these restrictions when stochastic volatility is allowed.

coefficients in multi-country VARs. If $\gamma_{ij}^{\text{DI}} = 0$, the coefficients on the lags of all country j variables for country i are set to very small values near zero. If $\gamma_{ij}^{\text{CSH}} = 0$, the coefficients on the lags of all country i variables for itself are to be concentrated at coefficients related to country j. Korobilis (2016) uses this prior in a multi-country forecasting exercise for bond yields of Eurozone countries and finds that it performs comparably with alternative shrinkage priors.

In terms of MCMC estimation, application of the DI restrictions is relatively straightforward, while application of the CSH restrictions is non-trivial, since we seek to use priors to push the model toward equality of matrices $B_{ii,\ell} = B_{jj,\ell}$ for $i \neq j$ and $\ell = 1, \ldots, p$. Koop and Korobilis (2016) provide a novel solution to the problem, but it still introduces prior dependence across equations, which also appears in the conditional posteriors, making it difficult to apply efficient algorithms as in Carriero, Clark, and Marcellino (2019) to estimate the model equation by equation. In addition, the priors of (7)-(8) involve tuning many hyperparameters. For example, if we specify hyper-priors to infer $\gamma_{ij}^{\text{DI}} \sim Bernoulli(\pi_{ij}), \gamma_{ii}^{\text{CSH}} \sim Bernoulli(\pi_{ii})$, and $\tau_{ij}^2 \sim Ga(a_1, a_{ij}), \xi_{ij}^2 \sim Ga(b_1, b_{ij})$, then we need to specify many more hyperparameters $(\pi_{ij}, \pi_{ii}, a_1, b_1, a_{ij}, b_{ij})$ related to those hyper-priors. Moreover, in contrast to the case of the scale mixtures of Normals priors we use, the theoretical properties are not known.

The next section introduces hierarchical shrinkage priors for multi-country VARs. Our priors make efficient MCMC algorithms easy to apply, without the need to tune many hyperparameters, and with the ability to push the model toward both CSH and DI restrictions.

3. Hierarchical Shrinkage in Multi-country VARs

The hierarchical shrinkage we consider is inspired by recent advances in the literature on scale mixtures of Normals priors and their successful applications in single-country Bayesian VARs. Since the seminal work by Carvalho, Polson, and Scott (2010), a variety of scale mixtures of Normals priors have been proposed in the literature. These priors include the Horseshoe prior (Carvalho, Polson, and Scott (2010)), the Normal-Gamma prior (Griffin and Brown (2010)), the Normal-Gamma-Gamma prior (Griffin and Brown (2017)), and several other alternatives. Compared to a conventional Normal prior, these prior distributions are spiked at the origin to provide severe shrinkage towards zero for the parameters of interest, while at the same time they also have heavy-tails to allow little shrinkage of, say, intercept terms in (2). In addition, these priors have computational advantages compared to the spike-and-slab prior (Carvalho, Polson, and Scott (2009)). Applications to single-country Bayesian VARs mostly focus on the Normal-Gamma prior; see, for instance, Huber and Feldkircher (2019) and Korobilis and Pettenuzzo (2019). Follett and Yu (2019) introduce the Horseshoe prior. These papers find that scale mixtures of Normals priors serve as competing alternatives to Minnesota priors, in terms of both forecasting and structural analysis.

For setting up priors on the parameters associated with model (2), let $\beta_i = vec([c_i, B_{ii,1}, \ldots, B_{ii,p}]'), i = 1, \ldots, N$, and let $\beta^{\text{CSH}} = (\beta_1, \ldots, \beta_N)'$ be the collection of coefficients related to CSH restrictions. Due to the forecasting focus, we consider further splitting β^{CSH} into three blocks:

$$\beta_c^{\text{CSH}} = (c_1', c_2', \cdots, c_N')',$$
$$\beta_{AR}^{\text{CSH}} = (\text{diag}(B_{11,1}), \text{diag}(B_{22,1}), \ldots, \text{diag}(B_{NN,1}), \ldots, \text{diag}(B_{11,p}), \ldots, \text{diag}(B_{NN,p})),$$
$$\beta_o^{\text{CSH}} = \beta^{\text{CSH}} \setminus \{\beta_c^{\text{CSH}}, \beta_{AR}^{\text{CSH}}\},$$

where β_o^{CSH} includes the set of parameters related to cross-variable lags that are in β^{CSH} and not related to the intercept terms and own lags. Similarly, let $\beta_{il}^* = vec([B_{i1,l}, \ldots, B_{ii-1,l}, B_{ii+1,l}, \ldots, B_{iN,l}]'), i = 1, \ldots, N, l = 1, \ldots, p$, and let $\beta^{\text{DI}_i} = (\beta_{i1}^*, \ldots, \beta_{ip}^*)'$ be the collections of coefficients related to DI restrictions in country i. Finally, define α as the free elements in A^{-1}. We have $N + 4$ blocks of coefficients: $\beta_c^{\text{CSH}}, \beta_{AR}^{\text{CSH}}, \beta_o^{\text{CSH}}, \beta^{\text{DI}_1}, \ldots, \beta^{\text{DI}_N}, \alpha$.

Using β as a generic notation for one block of coefficients, the hierarchical shrinkage we consider takes the form:

$$\beta_j \sim N(0, \lambda\omega_j), \quad \omega_j \sim \mathcal{F}, \tag{9}$$

where $j = 1, \ldots, \dim(\beta)$ and $\mathcal{F}$ denotes some pre-specified distribution for the local shrinkage parameter ω_j, which will be defined later. λ serves as a global shrinkage parameter, which can be specified as an additional hyper-prior to learn the values from the data. Taking β^{DI_1} as an example, because λ loads for all elements in β^{DI_1}, if $\lambda \to 0$, all β^{DI_1} are assumed to be identical (centered at zero), which implies that there is no dynamic interdependence for country 1. It is worth mentioning that, since both local and global shrinkage parameters are fully learned from the data, the hierarchical shrinkage approach offers more flexibility than the conventional Minnesota prior, in which all hyperparameters are fixed at some pre-specified values.

3.1. Choices of the priors

What remains is to specify $\mathcal{F}$ and hyper-priors on the global shrinkage parameters. Here we focus on three different scale mixtures of Normals priors, since they have been applied successfully in single-country Bayesian VARs.[6]

The first prior we consider is the Horseshoe prior:

$$\beta_j|\omega_j^2 \sim N(0, \omega_j^2), \quad \omega_j^2|\gamma_j^2 \sim \mathcal{G}(\frac{1}{2}, \gamma_j^2), \quad \gamma_j^2 \sim \mathcal{G}(\frac{1}{2}, \lambda), \tag{10}$$

where $\mathcal{G}$ denotes the Gamma distribution.[7] We use the parameterization of the Horseshoe prior as in

[6]Recently, there is a growing interest in the Dirichlet-Laplace prior; see Koop, et al. (2020) for an application. However, we do not consider it here, since the Dirichlet-Laplace prior is the scale mixture of the Laplace prior, which is not the main focus of this paper.

[7]We use the parameterization of the $\mathcal{G}(a, b)$ distribution with pdf given by $f(y) \propto y^{a-1} \exp(-by)$.

Armagan, Clyde, and Dunson (2011). It can be shown that the marginal distribution of ω_j^2 follows $C^+(0,1)$, where $C^+(0,1)$ denotes a half-Cauchy distribution on $\mathcal{R}^+$ with scale parameter 1, and λ serves as the global shrinkage parameter, which is the original parameterization in Carvalho, Polson, and Scott (2010) and used in Follett and Yu (2019). The prior (10) has computational advantages, since the conditional posteriors are conjugate (Makalic and Schmidt (2015)), making MCMC estimation straightforward. For the global shrinkage parameter λ, we also follow Armagan, Clyde, and Dunson (2011) and set $\lambda \sim C^+(0,1)$.

The second prior we consider is the Normal-Gamma prior:

$$\beta_j|\omega_j^2 \sim N(0,\omega_j^2), \;\; \omega_j^2 \sim \mathcal{G}(a^\omega, \frac{a^\omega \kappa^2}{2}). \tag{11}$$

The above, with a slightly different parameterization, was first introduced in Griffin and Brown (2010) and has recently been applied in single-country Bayesian VARs (Huber and Feldkircher (2019)) and time-varying parameter models (Bitto and Frühwirth-Schnatter (2019)). Using Monte-Carlo simulations, Bitto and Frühwirth-Schnatter (2019) find that a^ω controls the behavior in the neighborhood of the origin of the marginal prior distribution of $p(\beta_j)$ and κ^2 is the global shrinkage parameter. It can also be shown that (11) simplifies to Bayesian Lasso (Park and Casella (2008)) if $a^\omega = 1$ (Griffin and Brown (2010)). To infer hyperparameter values, we follow Bitto and Frühwirth-Schnatter (2019) to set $a^\omega \sim \mathcal{E}(b)$ and $\kappa^2 \sim \mathcal{G}(d_1, d_2)$, where $\mathcal{E}$ denotes the exponential distribution.

The final prior we consider is the Normal-Gamma-Gamma prior:

$$\beta_j|\tau_j^2, \lambda_j^2 \sim N\left(0, \phi\frac{\tau_j^2}{\lambda_j^2}\right), \;\; \tau_j^2 \sim \mathcal{G}(a,1), \;\; \lambda_j^2 \sim \mathcal{G}(c,1), \tag{12}$$

where $\phi = 2c/(a\kappa^2)$. The above, again with a slightly different parameterization, was first introduced in Griffin and Brown (2017). Cadonna, Frühwirth-Schnatter, and Knaus (2020) apply it to time-varying parameter models. They show that the specification in (12) is very general and nests some commonly used shrinkage priors. They also provide a comprehensive analysis of the properties of the Normal-Gamma-Gamma prior. It can be shown that a controls the behavior in the neighborhood of the origin and c controls the asymptotic tail behavior of the marginal prior distribution of $p(\beta_j)$. ϕ is the global shrinkage parameter. Using both simulated and real macroeconomic data, they find that the specification in (12) delivers relatively sparse solutions in time-varying parameter models. However, no attempts have been made so far to examine its performance for macroeconomic forecasting. Following Cadonna, Frühwirth-Schnatter, and Knaus (2020), we set $2a \sim \mathcal{B}(\alpha_a,\beta_a)$, $2c \sim \mathcal{B}(\alpha_c,\beta_c)$, and $\kappa^2|a,c \sim F(2a,2c)$ to learn these hyperparameter values, where $\mathcal{B}$ denotes the Beta distribution and F is the standard F distribution.

3.2. Comparisons of the priors

Theoretically, the scale mixtures of Normals priors are typically compared in terms of the concentration properties at the origin and the asymptotic tail behavior. Results for the Horseshoe prior and the Normal-Gamma-Gamma prior can be found in Carvalho, Polson, and Scott (2010) and Cadonna, Frühwirth-Schnatter, and Knaus (2020), respectively. However, no results are available for the Normal-Gamma prior. In the following theorem, we formally characterize the tail behavior and concentration properties for the Normal-Gamma prior.

Theorem 1. *Let $\beta_j \sim \mathcal{NG}(a^\omega, \kappa^2)$, where $\mathcal{NG}$ is the Normal-Gamma prior parameterized as in* (11). *Then, the marginal density $\pi_{NG}(\beta_j)$ satisfies the following:*

- *Concentration properties: As $\left|\beta_j\right| \to 0$, we have*
  1. *if $a^\omega > \frac{1}{2}$, $\pi_{NG}(\beta_j) = O(1)$;*
  2. *if $0 < a^\omega < \frac{1}{2}$, $\pi_{NG}(\beta_j) = O\Big(\frac{1}{|\beta_j|^{\frac{1}{2}-a^\omega}}\Big)$;*
  3. *if $a^\omega = \frac{1}{2}$, $\pi_{NG}(\beta_j) = O\Big(\frac{1}{\log\left(|\beta_j|\right)}\Big)$;*
- *Asymptotic tail behavior: As $\left|\beta_j\right| \to \infty$, we have $\pi_{NG}(\beta_j) = O\left(\frac{|\beta_j|^{a^\omega-1}}{\exp\left(\sqrt{a^\omega\kappa^2}|\beta_j|\right)}\right)$.*

Proof. See Appendix A. □

The results for Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma priors in terms of both asymptotic tail behavior and concentration properties are summarized in Table 1. Clearly, Normal-Gamma and Normal-Gamma-Gamma priors share similar concentration properties, possessing unbounded density near the origin if either $0 < a^\omega < \frac{1}{2}$ or $0 < a < \frac{1}{2}$. Both priors diverge to infinity with a polynomial order, much faster than the Horseshoe prior (with a logarithmic order). For the tail behavior, it follows from straightforward calculation that $\lim_{|\beta_j|\to\infty} \pi_{NG}(\beta)/\beta^{-2} = 0$ and $\lim_{|\beta_j|\to\infty} \pi_{NGG}(\beta)/\beta^{-2} = \infty$ if $0 < c < \frac{1}{2}$, which implies that the Normal-Gamma prior has lighter tails than the Horseshoe prior, but the Normal-Gamma-Gamma prior has heavier tails than the Horseshoe prior. The Normal-Gamma-Gamma prior is the only one that can achieve a polynomial rate of convergence in both the tails and the origin. It extends the Normal-Gamma prior by having heavier tails. Compared to the Horseshoe prior, it puts more probability mass at the origin and offers more flexibility in modeling tails by the hyperparameter c.

Table 1: Tail behavior and concentration around zero for Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma priors

	Tail Decay	Concentration at zero
Horseshoe	$O\Big(\frac{1}{\beta_j^2}\Big)$	$O\Big(\log\big(\frac{1}{\lvert\beta_j\rvert}\big)\Big)$
Normal-Gamma	$O\Bigg(\frac{\lvert\beta_j\rvert^{a^\omega-1}}{\exp\big(\sqrt{a^\omega\kappa^2}\lvert\beta_j\rvert\big)}\Bigg)$	$O(1)$ if $a^\omega > \frac{1}{2}$
		$O\Big(\frac{1}{\lvert\beta_j\rvert^{\frac{1}{2}-a^\omega}}\Big)$ if $0 < a^\omega < \frac{1}{2}$
		$O\Big(\frac{1}{\log(\lvert\beta_j\rvert)}\Big)$ if $a^\omega = \frac{1}{2}$
Normal-Gamma-Gamma	$O\Big(\frac{1}{\beta_j^{2c+1}}\Big)$	$O(1)$ if $a > \frac{1}{2}$
		$O\Big(\frac{1}{\lvert\beta_j\rvert^{1-2a}}\Big)$ if $0 < a < \frac{1}{2}$
		$O\Big(\frac{1}{\log(\lvert\beta_j\rvert)}\Big)$ if $a = \frac{1}{2}$

Empirically, Cadonna, Frühwirth-Schnatter, and Knaus (2020) use Euro area macroeconomic data and find that the Normal-Gamma-Gamma prior achieves more sparse parameter estimates than other shrinkage priors in a time-varying parameter model framework. However, as pointed out in Giannone, Lenza, and Primiceri (2021), sparsity does not necessarily imply good forecasting performance. As we shall see, heavy tails in the prior distributions of the coefficients is an important feature to obtain better forecasting performance in multi-country VARs, since the Horseshoe prior forecasts well in many cases. The extension to the Normal-Gamma-Gamma prior is less useful. The light-tailed Normal-Gamma prior is useful for output growth forecasts in some cases, but it is outperformed by Horseshoe and Normal-Gamma-Gamma priors for inflation and interest rate forecasts.

3.3. Other competing models

In addition to the multi-country VARs mentioned above, we also consider several alternative models commonly used in macroeconomic forecasting. These models include country-specific VARs with SV in which priors are specified as either Minnesota-type or hierarchical Normal-Gamma as in Chan (2021);[8] country-specific factor-augmented VAR (FAVAR) models with SV; global VAR (GVAR) with SV; and multi-country VAR-SV with factor shrinkage as in Canova and Ciccarelli (2009). We use the country-specific VAR-SV with Minnesota prior as the benchmark, as it is the most commonly used model in the macroeconomic forecasting literature. A description of all the models under comparison is provided in Table 2. More details on the specification of the various models and associated priors can be found in the Appendix.

[8] We use a specification from a published paper as a competitor. In Section 6.4, we compare forecast performance to country-specific VAR-SV with the hierarchical prior specification proposed in this paper. We find it delivers more accurate forecasts than Chan (2021), but it is still outperformed by the multi-country VAR-SV with hierarchical shrinkage.

Table 2: List of competing models

Model	Description
CVAR	country-specific VAR(p) with Minnesota prior
CVAR-H	country-specific VAR(p) with hierarchical shrinkage as in Chan (2021)
CFAVAR	country-specific factor-augmented VAR(p), with factors extracted from foreign variables
GVAR	Global VAR(p)
CC	parameters are assumed to follow an exact factor structure, as in Canova and Ciccarelli (2009)
MIN	priors are Minnesota-type as in Angelini, et al. (2019)
SSSS	stochastic specification search and selection prior as in Korobilis (2016)
HS	hierarchical shrinkage with Horseshoe prior
NG	hierarchical shrinkage with Normal-Gamma prior
NGG	hierarchical shrinkage with Normal-Gamma-Gamma prior

Note: All the models include SV.

4. Estimation Algorithms

4.1. Estimation outline

We estimate all of the models listed in Table 2 using Markov Chain Monte Carlo (MCMC) methods. All of our estimates are based on 30,000 posterior draws, with the first 10,000 discarded and the remaining 20,000 post-burn-in draws retained. This section provides a brief overview of our methods. The Appendix and the studies cited below provide more details on algorithms and priors.

For country-specific VARs with SV and factor-augmented VARs with SV, we use the non-conjugate Minnesota-type prior. The Gibbs sampling details of the country-specific VAR with SV are provided in Carriero, Clark, and Marcellino (2019). The intercept and autoregressive coefficients are estimated by the corrected triangular algorithm proposed in Carriero, et al. (2022).[9] Stochastic volatility is estimated with the algorithm in Del Negro and Primiceri (2015). For free elements in A, we use the algorithm as in Cogley and Sargent (2005). The Gibbs sampler used in country-specific VARs with SV can be easily extended to allow for augmented factors extracted from foreign variables. For the GVAR, we use a Minnesota-type prior similar to Huber (2016); details of the algorithms are provided there. For the CVAR-H, CFAVAR, and GVAR specifications, we follow Chan (2021) and put hyper-priors on the overall shrinkage parameters related to own lags and cross-variable lags. An additional step is needed to update these parameters. In the case of the CVAR-H model, we use the default setting as in Chan (2021), and sampling details can be found there.

To estimate the multi-country VAR with SV and the factor shrinkage approach (the CC model), we use

[9] A summary of the corrected triangular algorithm is provided in Appendix C.4.

an exact factorization as in Canova, Ciccarelli, and Ortega (2007) and Korobilis (2016). Algorithms are provided in these papers. SV can be easily added to this model and estimated similarly as in the country-specific case. In the case of the MIN specification that features a Minnesota-type prior, the algorithm can be obtained similarly, but for the Minnesota-type prior we use a specification similar to Angelini, et al. (2019). We also put a hyper-prior on the overall shrinkage parameter related to coefficients on lagged foreign variables. Finally, for the three hierarchical shrinkage approaches (HS, NG, and NGG), the Gibbs samplers are again very similar to those in other models, but additional blocks of sampling are needed to update the hyperparameters. In particular, the details of the NG and NGG priors can be adapted as in Bitto and Frühwirth-Schnatter (2019) and Cadonna, Frühwirth-Schnatter, and Knaus (2020), respectively. For the HS prior, since we use a different parameterization, algorithms in Follett and Yu (2019) cannot be directly applied, but sampling schemes in Armagan, Clyde, and Dunson (2011) for univariate regression models can be extended. We provide a summary of the MCMC algorithms below.

Algorithm 1: MCMC inference for multi-country VARs with SV

Step 1: initialization;

Step 2: **for** $i = 1, \ldots, NG$ **do**

  Use the corrected triangular algorithm in Carriero, et al. (2022) to obtain posterior draws from VAR mean coefficients (intercepts and autoregressive coefficients);

end

Step 3: Use the algorithm in Cogley and Sargent (2005) to update the free elements in A;

Step 4: Update the hyperparameters in prior error covariance matrices, with conditional posteriors that depend on prior choices, which can found in Appendix B.2;

Step 5: Use the algorithm in Del Negro and Primiceri (2015) to update the volatilities h_t and error variance of innovations Φ^h.

4.2. Computational efficiency

In this section, we briefly compare the computational efficiency of the MCMC algorithms for the multi-country VARs (note that we omit the CC specification, which is covered elsewhere in the literature). To assess the efficiency of the algorithms, we compute the potential scale reduction factors (PSRFs) detailed in Brooks and Gelman (1998). A value of the PSRFs below 1.1 is generally taken as an indication that the chain has satisfactory mixing properties. In Table 3, we report average PSRFs of parameters needed to construct forecasts and obtain predictive distributions: vec(B), α, and vech(Φ). We use the data set as in our forecasting exercises (21 variables) and all models include 4 lags. As is clear, our algorithms show satisfactory mixing and convergence properties.

Table 3: Mixing and convergence statistics (PSRFs) for multi-country VARs with 21 variables

	MIN	SSSS	HS	NG	NGG
vec(B)	1.005	1.002	1.002	1.002	1.008
α	1.004	1.004	1.002	1.002	1.004
vech(Φ)	1.061	1.040	1.046	1.077	1.069

Notes: The lags are set to 4 for all models except SSSS, in which we use 1 lag, to match the specification of our forecasting application.

As we discussed above, to handle CSH restrictions, the SSSS prior introduces dependence across equations, which prevents the use of an efficient sampling algorithm for the VAR's coefficients. Table 4 shows the computational time (in seconds) necessary to produce 10,000 draws from the posteriors of multi-country VARs including 21 variables. For this time comparison, all models include only 1 lag. Clearly, the specification with the SSSS prior takes much longer to estimate, roughly 5 times slower than the other specifications. In forecasting, computations can be extremely burdensome as estimation has to be done many times. Interestingly, the added blocks of sampling for hyperparameters in our proposed HS, NG, and NGG methods have very small additional computational costs compared to the Minnesota-type prior of the MIN specification.

Table 4: Time (in seconds) taken to obtain 10,000 posterior draws for various multi-country VARs with 21 variables and 1 lag

MIN	SSSS	HS	NG	NGG
393	2077	367	324	427

5. Data and Forecast Evaluation

5.1. Data

We examine the forecast performance of the various specifications using a data set for G7 countries: USA, UK, Germany (DEU), France (FRA), Italy (ITA), Japan (JPN), and Canada (CAN). In brief, we build a 3-variable data set for each country at a quarterly frequency, with a sample period of 1973Q1-2019Q4. The variables consist of real GDP growth, CPI inflation, and a short-term interest rate (the 3-month government bill rate). Table 5 presents the details of the data set along with the transformations of the variables and the corresponding sources.[10] Note that, like most other multi-country studies, we do not consider real-time data and use data from the last available vintage, owing to the lack of availability of real-time data for all seven countries.

[10]We obtain data from different sources due to data availability. In particular, the OECD provides data on real GDP growth for Germany before 1991, and the Global Financial Database (GFD) provides very long coverage for interest rate data in many countries.

Table 5: Data description and variable transformation

Variable	Data source	Transformation
Real GDP growth	OECD	$4y_t$
CPI inflation	FRED	$400\log(y_t/y_{t-1})$
Interest rate	GFD	None

Notes: Because the OECD reports GDP growth as quarterly percent changes, we multiply the source data by 4 to obtain an approximate annual rate. FRED refers to the database maintained by the Federal Reserve Bank of St. Louis. CPI inflation is measured with the annualized quarterly percent change in the quarterly average level of the monthly CPI. GFD refers to the Global Financial Database, from which we obtain the 3-month government bill rate and form the quarterly series as the average of values for the months of each quarter.

5.2. *Forecast evaluation*

We consider both point and density forecasts at horizons up to 12 steps (three years) ahead. Parameter estimation and out-of-sample forecasting are done recursively, using an expanding window of data for model estimation. The initial estimation sample runs from 1973Q1 to 1994Q4, the first available forecast is for 1995Q1, and forecasts are generated up to 12 quarters ahead. Our last estimation sample runs from 1973Q1 to 2016Q4, yielding forecasts from 2017Q1 to 2019Q4.

For all the models considered here, the full distribution of the forecasts is not available in closed form, and a simulation algorithm is required. At each post-burn-in draw, we compute the implied path of $\hat{y}^{(j)}_{t+h}$ to generate a total of 20,000 draws from the predictive distribution.

Each point forecast is measured as the median of the predictive density. We evaluate them in terms of root mean squared forecast error (RMSFE). Letting $\hat{y}_{t+h}(M)$ be the forecast of the (scalar, for simplicity of notation here) target variable y_{t+h} made by model M and letting P be the total number of generated forecasts, the RMSFE made by model M for horizon h is

$$\text{RMSFE}^M_h = \sqrt{\frac{1}{P}\sum(\hat{y}_{t+h}(M) - y_{t+h})^2}. \tag{13}$$

In the case of the density forecasts, we use the continuous ranked probability score (CRPS) proposed by Gneiting and Raftery (2007), which is less sensitive to outliers than other density evaluation measures, such as the log score. The CRPS metric for the each variable at time t for horizon h is defined as

$$\text{CRPS}_t(F, y^o_{t+h}) = \mathbb{E}_F\left|y^d_{t+h} - y^o_{t+h}\right| - \frac{1}{2}\mathbb{E}_F\left|y^d_{t+h} - y^{dd}_{t+h}\right|, \tag{14}$$

where F denotes the cumulative distribution function associated with the predictive density f, y^o_{t+h} denotes the observed value, and y^d_{t+h}, y^{dd}_{t+h} are independent draws from the predictive posterior distribution. Following Smith and Vahey (2016), we compute (14) by numerical integration methods, which are shown to be more accurate and efficient. It is also worth mentioning that the lower the value of the CRPS, the more accurate the predictive density is.

Finally, to provide a statistical comparison of predictive accuracy, we apply the Diebold and Mariano (1995) (DM) test for equal forecast accuracy. Yet, our models are nested in many cases. It is well known that the DM test for nested models is undersized, and the results can be viewed as conservative for equal forecast accuracy in finite samples. We follow Coroneo and Iacone (2020) to apply fixed-smoothing asymptotics for the DM test, which is shown to deliver predictive accuracy tests that are correctly sized even when the number of out-of-sample observations are small.

6. Empirical Results

6.1. Overall forecast performance

We first provide a summary of the forecast evaluation exercise in Table 6. As we have 7 countries and forecasts are generated from 1 to 12 steps ahead, for each variable we have 84 forecasts from each model. The table reports the number of cases in which each model is best, for all horizons, short horizons ($h \leqslant 6$), and long horizons ($h > 6$).

Based on Table 6, the results can be summarized as follows. First, our proposed hierarchical shrinkage in multi-country VARs — used in the HS, NG, and NGG specifications — is quite helpful. This is particularly true with the HS prior. It has the most wins for inflation in terms of both point and density forecasts, and it is the best performing model for output growth and the interest rate in terms of density forecasts. More specifically, for output growth, the HS prior is the best in 31 (out of 84) cases in terms of density forecasts, compared to 15 cases for the second-best performing model, which is the multi-country VAR with the NG prior. For inflation, the HS specification performs the best in more than half (44) of the cases in terms of point forecasts and exactly half of the cases in terms of density forecasts. The benefits are also more evident at long horizons, with 29 wins in point forecasts and 27 wins in density forecasts. For the interest rate, the HS prior also has the most wins (22 cases) for density forecasts, compared to 16 cases obtained from the factor shrinkage approach of the CC specification. Second, the NG specification is the best in nearly half (40) of the cases for output growth in terms of point forecasts, compared to 19 cases from the SSSS prior. The CC specification has the most wins for the interest rate in terms of point forecasts. However, the NG and CC specifications do not forecast well for other variables. For instance, for inflation at long horizons, the NG and CC specifications are never selected as best. SSSS never becomes the best for density forecasts of the interest rate. Third, including information across countries is very useful particularly for output growth; the single-country CVAR and CVAR-H specifications are never the best in point or density forecasts of output. Although the baseline single-country CVAR has the second best forecast performance for inflation, with 13 wins in terms of point forecasts and 10 wins for density forecasts, it is clearly outperformed by the multi-country HS specification. Regarding the CFAVAR and GVAR specifications, they perform better for inflation than the other variables but also fall short of the HS approach. The single-country CVAR-H specification that imposes hierarchical shrinkage works better for the interest rate than for output growth or inflation, but still falls short of other models, including the multi-country HS specification.

Overall, the usefulness of hierarchical shrinkage for multi-country VARs in forecasting key macroeconomic variables emerges rather clearly. In general, the HS prior is better than the other two scale mixtures of Normals priors. As discussed in Section 3, with more hyperparameters controlling both origins and tails, the NGG prior is theoretically more flexible than the HS prior, which also provides a heavy-tailed extension of the NG prior. However, the theoretical advantages do not necessarily transfer to better forecasting performance, as it only ranks first (tied) in the case of density forecasts of the interest rate at short horizons.

The summary results should be interpreted with care, as they are based on deterministic comparisons (i.e., the best model could be not statistically better than the second-best model). They also ignore the cross-country differences and the potential differences of model performance over time. Yet, they provide a broad overview of the models' performance. More detailed results and statistical comparisons are presented in the next subsection.

Table 6: Summary statistics: number of cases when one model becomes the best

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	point	density	point	density	point	density	**Inflation**	point	density	point	density	point	density
CVAR	0	0	0	0	0	0	CVAR	13	10	7	7	6	3
CVAR-H	0	0	0	0	0	0	CVAR-H	0	2	0	2	0	0
CFAVAR	1	3	0	0	1	3	CFAVAR	8	6	3	2	5	4
GVAR	0	0	0	0	0	0	GVAR	3	10	2	3	1	7
CC	4	1	2	1	2	0	CC	0	0	0	0	0	0
MIN	5	7	1	2	4	5	MIN	8	9	8	8	0	1
SSSS	19	11	12	11	7	0	SSSS	4	3	4	3	0	0
HS	9	31	4	9	5	22	HS	44	42	15	15	29	27
NG	40	15	21	12	19	3	NG	1	2	1	2	0	0
NGG	6	16	2	7	4	9	NGG	3	0	2	0	1	0
Interest rate	point	density	point	density	point	density							
CVAR	6	10	3	4	3	6							
CVAR-H	6	13	6	9	0	4							
CFAVAR	0	0	0	0	0	0							
GVAR	0	1	0	1	0	0							
CC	30	16	11	3	19	13							
MIN	1	1	1	1	0	0							
SSSS	8	0	2	0	6	0							
HS	19	22	7	8	12	14							
NG	10	11	8	7	2	4							
NGG	4	10	4	9	0	1							

Notes: See Table 2 for a list of models and Section 5.2 for the evaluation criteria.

6.2. Forecast evaluation: Cross-country differences

Building on the previous section's overview of the forecast performance of the various model specifications, we turn now to a more quantitative assessment of forecast accuracy across models and countries. To this end, Tables 7-9 report relative RMSFEs and CRPSs for all G7 countries at selected horizons, $h = 1, 4, 8, 12$. Entries shaded in gray indicate the best performing model. RMSFEs and CRPSs in levels from the benchmark model are reported in the appendix's Table D.10.

Consider first the results for output growth. In general, the best performing specifications are the multi-country VARs with scale mixtures of Normals priors (i.e., one of HS, NG, or NGG). In all but a few cases, these specifications improve on the point and density forecast accuracy of the CVAR benchmark. In general, hierarchical shrinkage of multi-country VARs, in particular with the HS prior, is rather a safe option for forecasters, since it provides gains for both point and density forecasts in most of the cases. The other specifications do not fare as well in improving on the accuracy of the benchmark. One of the better alternatives is the SSSS specification, which is best for a few countries at short horizons, although in some other cases it is fairly strongly beaten by both other approaches (i.e., its performance is somewhat uneven). The MIN specification — a multi-country VAR estimated with a Minnesota-type prior — is only best in a few instances, all of them long-horizon forecasts for the USA. Similarly, the CC (factor-based shrinkage of coefficients) and CFAVAR (factor-augmented single-country models) are selected as best for no more than a few country/horizon/type of forecast combinations. Perhaps not surprisingly, the accuracy of the CVAR-H specification (single-country with hierarchical shrinkage) is relatively similar to the CVAR baseline, sometimes a little better and sometimes a little worse.

Moving to the inflation forecasts of Table 8, some commonalities and some different stories are both evident. First, the multi-country HS specification continues to provide the best forecast in many cases, particularly for Canada, France, Italy, and Japan at longer horizons, as well as for density forecasts for the USA. Second, other competing specifications, including CFAVAR, GVAR, MIN, and SSSS, are occasionally the best model, but they are relatively less accurate in other cases. For example, the SSSS specification is the best for the USA at the 1-step-ahead and 4-steps-ahead horizons in terms of point forecasts and 4-steps-ahead horizon in terms of density forecasts, but it does not provide gains to forecasts for Italy. Third, results are somewhat different for the UK, perhaps because historical inflation in the UK is rather different, with stronger peaks in the 1970s and a volatile period around the Black Wednesday crisis. In the case of the UK, the best-performing forecasting model is the benchmark specification, with RMSFE and CRPS ratios that exceed 1 in all but one case.

Turning to the interest rate forecasts, which are presented in Table 9, we again see similarities as well as some different patterns. Sorting through differences across countries, the multi-country HS specification continues to perform relatively well. For most, although not all, countries, forecasts from this model are more accurate than the benchmark, by margins as large as 32 percent. The other two multi-country scale mixtures of Normals priors — NG and NGG — don't offer any clear advantages over the HS specification, sometimes slightly to modestly improving accuracy and other times reducing accuracy (relative to the HS prior). Of the other multi-country VAR specifications, the CC model performs better in forecasting interest rates than output growth and inflation. For a few country/horizon combinations, the CC model is most accurate, whereas for some others, it is notably less accurate than the CVAR benchmark. The performance of the SSSS specification is also uneven, often much less accurate than the benchmark (e.g., for Canada) but occasionally more accurate (e.g., 4- and 8-steps-ahead forecasts for Germany). The performance of the GVAR is also inconsistent.

Table 7: Out-of-sample output growth forecast performance: RMSFE and CRPS ratios in terms of CVAR benchmark, selected horizons

RMSFE	Canada									Germany								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
h = 1	0.993	0.903	0.940	1.008	1.000	0.845	0.927	0.865	0.861	0.981	0.891	0.937	1.084	0.981	0.853	0.972	0.899	0.921
h = 4	1.007*	1.009	0.986*	1.117	0.991	0.992	0.945	0.954	0.947	0.998	0.987	0.990	1.141	0.982	0.977	0.996	0.994	0.995
h = 8	1.006	0.993	1.002	1.237	1.003	1.032	0.976	0.975	0.982	0.997	0.996	1.001	1.193*	1.002	1.019	0.985	0.974	0.983
h = 12	1.000	0.988	0.992	1.268	1.013	1.031	0.988	1.003	0.994	0.999	0.996	1.001	1.186	0.996	1.025	0.974	0.977	0.976
	France									Italy								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
h = 1	0.989	0.962	0.971	1.038	1.024	0.954	0.971	0.934	0.929	0.981	0.980	1.008	0.959	1.075	0.981	1.021	0.984	1.000
h = 4	0.991	0.991	0.996	1.190	0.977	1.032	0.933	0.908	0.920	0.999	0.986	1.017	1.070	1.007	1.010	0.957	0.935	0.952
h = 8	0.985*	1.002	0.996	1.290	0.993	1.117	0.931	0.919	0.924	1.004	1.002	1.027*	1.020	1.016	1.022*	0.961	0.953	0.955
h = 12	0.983	1.001	0.994	1.374	1.001	1.143*	0.932	0.921*	0.940	1.008	0.998	1.027*	0.967	1.020	1.003	0.971	0.980	0.980
	Japan									UK								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
h = 1	1.005	0.983	0.993	1.128	1.014	0.948*	0.983	0.966	0.971	1.019	1.018	1.009	0.924	1.136	0.954	1.000	1.051	1.015
h = 4	0.991	0.986	0.996	1.122	0.988	0.978	0.978	0.954	0.966	1.003	1.002	1.005	1.075	0.974	0.943	0.949	0.930	0.936
h = 8	0.994	0.995	0.996	1.126	0.999	0.979	0.972	0.963	0.958	1.013*	1.006	1.002	1.099	0.981	0.959	0.972	0.965	0.961
h = 12	0.990	0.991	0.993	1.102	0.990	1.010	0.969*	0.960	0.956	1.007	1.002	1.001	1.206	1.001	0.959	0.989	1.003	0.988
	USA																	
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG									
h = 1	0.989	0.972	1.010	1.080	1.027	0.921	0.977	0.936	0.947									
h = 4	0.985	1.016	1.025	1.223	0.968	0.965	0.973	0.966	0.965									
h = 8	0.994	1.016	1.020	1.386	0.962	0.962	0.984	1.013	0.992									
h = 12	0.986	1.027*	1.035	1.499*	0.974	0.976	0.986	1.001	0.990									
CRPS	Canada									Germany								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
h = 1	0.994	0.935	0.953	1.053	0.990	0.880	0.943	0.902	0.894	0.997	0.913	0.961	1.169*	0.972	0.880*	0.974	0.933	0.947
h = 4	1.009	1.011	0.978*	1.167	0.989	0.971	0.943	0.956	0.951	1.008	0.989	0.991	1.332*	0.982	0.969	1.002	1.005	1.004
h = 8	1.005	0.994	0.996	1.314	1.020	1.017	0.981	0.986	0.993	1.008	0.997	0.998	1.432*	1.000	1.011	0.977	0.982	0.980
h = 12	1.005	0.989	0.989	1.390*	1.020	1.054	0.990	1.016	0.997	1.009	1.000	0.998	1.480*	0.995	1.043	0.980	1.002	0.987
	France									Italy								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
h = 1	0.994	0.968	0.980	1.065	1.002	0.947	0.964	0.946	0.932	0.994	0.994	1.023	0.978	1.062	0.995	1.002	1.005	0.996
h = 4	0.984	0.997	0.990	1.233	0.968	1.007	0.922	0.907	0.915	1.000	0.996	1.025	1.098	1.013	1.003	0.958	0.944	0.957
h = 8	0.974*	1.006	0.996	1.383	0.988	1.089*	0.915	0.918	0.916	0.998	1.013	1.034	1.121	1.019	1.029	0.951	0.961	0.952
h = 12	0.973*	0.997	0.996	1.493	0.997	1.133*	0.924	0.921	0.929	1.007	0.992	1.030	1.150	1.008	1.039	0.954	0.976	0.963
	Japan									UK								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
h = 1	0.998	0.983	0.993	1.125	1.010	0.955*	0.979	0.969	0.960	1.031	1.020	1.008	0.976	1.114	0.970	1.000	1.055	1.018
h = 4	0.991	0.986	0.996	1.251*	0.990	0.978	0.973	0.952	0.958	1.008	1.005	1.003	1.149	0.971	0.945	0.944	0.937	0.947
h = 8	1.012	0.993	0.997	1.316*	1.009	0.973	0.974	0.972	0.959	1.013	1.004	0.998	1.245	0.986	0.977	0.968	0.975	0.976
h = 12	1.008	0.992	0.994	1.294*	0.995	0.987	0.970	0.975	0.959	1.013	0.999	1.004	1.384*	1.009	1.030	0.990	1.030	1.007
	USA																	
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG									
h = 1	0.991	0.987	1.009	1.127*	1.004	0.945	0.965	0.942	0.944									
h = 4	0.989	1.024	1.024	1.319*	0.973	0.976	0.973	0.978	0.976									
h = 8	0.987	1.024	1.022	1.506*	0.958	0.970	0.978	1.007	0.992									
h = 12	0.990	1.033	1.035	1.628*	0.971	1.020	0.967	0.997	0.977									

Notes: The models are detailed in Table 2. For each specification, the upper panel presents the ratios of RMSFEs relative to the CVAR benchmark. The lower panel presents the ratios of CRPSs relative to the CVAR benchmark. Values below 1 indicate the model outperforms the benchmark and vice versa. Gray shading indicates the best performing model. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed-smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 8: Out-of-sample inflation forecast performance: RMSFE and CRPS ratios in terms of CVAR benchmark, selected horizons

RMSFE

	Canada									Germany								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	1.000	1.005	1.019	1.037	0.979	0.998	0.978	0.962	0.965	0.999	0.996	0.988	1.136*	0.996	1.062	1.004	1.023	1.020
$h=4$	1.000	1.019	1.037	1.063	0.931	0.996	0.937	0.965	0.956	1.010	0.950	0.955	1.237*	0.947	1.010	0.968	1.004	0.982
$h=8$	1.001	1.053	0.990	1.134	0.934	1.086	0.901	1.024	0.934	1.027	0.914	0.925	1.580*	0.929*	1.020	0.934	0.985	0.962
$h=12$	0.978	1.029	0.955	1.228	0.932	1.185	0.845	0.998	0.887	1.017	0.900	0.899	1.937	0.934	1.091	0.897*	0.964	0.928

	France									Italy								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	1.003	0.999	1.006	1.039	0.998	1.048	0.996	1.019	0.999	1.004	0.998	1.014	1.040	1.005	1.021	1.010	1.030	1.002
$h=4$	1.017	1.029	1.011	1.029	0.949	0.988	0.938	0.996	0.955	1.005	1.047	1.036	1.122	1.014	1.048	0.979	1.059	0.989
$h=8$	1.009	1.058	1.011	1.252	1.006	1.041	0.932	1.058	0.971	0.980	1.081	1.049	1.261	1.090	1.177*	0.914	1.151	0.970
$h=12$	1.017	1.082	0.990	1.535*	1.032	1.091	0.912	1.151	0.987	0.971	1.090	1.017	1.386	1.117	1.318	0.852	1.199*	0.952

	Japan									UK								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	0.997	0.989	0.985	1.138*	0.964	1.059	0.980	0.998	0.978	1.026	1.024	1.015	1.005	1.006	1.079	1.035	1.121	1.092
$h=4$	1.004	1.026	0.978	1.115*	0.973	1.045	0.969	1.020	0.985	1.051	1.126	1.078	1.131	1.116*	1.192	1.130	1.363*	1.246
$h=8$	1.008	1.026	0.968	1.256*	0.960	1.055*	0.947	1.007	0.962	1.058	1.170	1.053	1.235	1.196	1.292	1.088	1.407*	1.179
$h=12$	1.003	1.000	0.963	1.369*	0.952	1.042	0.943	1.006	0.954	1.038	1.190	1.008	1.285	1.276	1.444*	1.036	1.489	1.151

	USA								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	1.016	0.970	0.976	1.060	0.955	0.953	0.967	0.973	0.961
$h=4$	1.021	0.993	0.960	1.139*	0.975	0.943	0.973	1.010	0.982
$h=8$	1.021	1.045	0.976	1.297*	1.024	1.018	0.980	1.040	0.993
$h=12$	0.999	1.048	0.980	1.500*	1.018	1.021	0.967	1.017	0.966

CRPS

	Canada									Germany								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	1.002	1.010	1.014	1.040	0.979	0.995	0.987	0.972	0.972	0.989	0.991	0.980	1.122*	0.982	1.057*	0.999	1.010	1.011
$h=4$	1.003	0.991	0.993	1.088*	0.911*	0.990	0.933	0.928	0.939	1.002	0.953	0.954	1.214*	0.944	1.029	0.968	0.995	0.981
$h=8$	1.019	1.001	0.945	1.186*	0.902*	1.086	0.886	0.978	0.910	1.016	0.900	0.916	1.534*	0.926*	1.044	0.937	0.974	0.955
$h=12$	1.008	0.987	0.916*	1.304*	0.889*	1.186	0.849*	0.956	0.873*	1.020*	0.899*	0.893*	1.925*	0.932*	1.183	0.906	0.966	0.931

	France									Italy								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	1.002	1.003	1.009	1.041	0.986	1.047	0.986	1.014	0.988	0.992	1.018	1.006	1.055*	1.013	1.030	1.019	1.069	1.023
$h=4$	1.020	1.024	1.010	1.010	0.944*	1.006	0.930*	0.996	0.944	1.006	1.068	1.035	1.166*	1.004	1.049	0.999	1.074	1.009
$h=8$	1.011	1.039	0.996	1.201	0.987	1.102*	0.913	1.017	0.942	0.989	1.083	1.045	1.279*	1.073	1.202*	0.919	1.115	0.961
$h=12$	1.018	1.053	0.955*	1.469*	0.995	1.251*	0.888	1.101	0.950	0.980	1.083	0.995	1.380*	1.077	1.388*	0.847	1.142*	0.917

	Japan									UK								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	0.994	0.992	0.985	1.146*	0.978	1.053	0.970	1.012	0.975	1.036	1.037	1.012	1.032	1.046	1.110*	1.030	1.114	1.073
$h=4$	1.011	1.024	0.980	1.208*	0.998	1.070*	0.961	1.026	0.976	1.043*	1.091	1.040	1.142*	1.128	1.200*	1.077	1.302*	1.170*
$h=8$	1.017	1.034	0.968	1.337*	0.989	1.098*	0.941	1.021	0.959	1.069	1.117	1.017	1.266*	1.181	1.319*	1.052	1.334*	1.128
$h=12$	1.004	1.016	0.966	1.448*	0.983	1.156*	0.940	1.032	0.955	1.019	1.141	0.986	1.359*	1.187	1.489*	1.027	1.377*	1.127

	USA								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h=1$	1.029	0.988	0.984	1.098	0.987	0.992	0.993	1.029	0.996
$h=4$	1.047*	1.025	0.963	1.211*	0.993	0.957	0.971	1.039	0.990
$h=8$	1.048	1.052	0.962	1.379*	1.008	0.995	0.968	1.055	0.993
$h=12$	1.037	1.048	0.954	1.661*	0.994	1.059	0.940	1.022	0.960

Notes: The models are detailed in Table 2. For each specification, the upper panel presents the ratios of RMSFEs relative to the CVAR benchmark. The lower panel presents the ratios of CRPSs relative to the CVAR benchmark. Values below 1 indicate the model outperforms the benchmark and vice versa. Gray shading indicates the best performing model. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed-smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 9: Out-of-sample interest rate forecast performance: RMSFE and CRPS ratios in terms of CVAR benchmark, selected horizons

RMSFE

Canada	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.993	0.994	1.044	1.036	1.084	1.210	1.021	1.075	1.044
$h = 4$	0.968	1.025	1.042	0.938	1.141	1.382	0.987	1.044	1.038
$h = 8$	0.961	1.042	1.040	0.815	1.147	1.708*	0.912*	0.985	0.970
$h = 12$	0.956	1.016	1.035	0.834	1.087	2.147*	0.830*	0.925	0.875*

Germany	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.011	0.939	0.939	1.065	0.980	1.005	0.946	0.947	0.929
$h = 4$	1.059	0.853*	0.826*	1.018	0.851*	0.839	0.791*	0.739*	0.774*
$h = 8$	1.080*	0.802*	0.777*	1.094	0.829*	0.953	0.727*	0.736*	0.775*
$h = 12$	1.067*	0.776*	0.763*	1.213	0.863*	1.201	0.700*	0.776*	0.789*

France	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.954	0.937	0.922	0.931	0.906	1.087	0.932	0.945	0.896
$h = 4$	0.983	0.929	0.881	0.822	0.863*	1.183	0.856	0.902	0.833
$h = 8$	0.989	0.928	0.878	0.767	0.886*	1.435	0.783	0.832*	0.778*
$h = 12$	0.986	0.966	0.900	0.883	0.966	1.918	0.777	0.908*	0.821*

Italy	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.991	0.965	0.971	1.044	0.942	1.136	0.948	1.063	0.959
$h = 4$	0.988	0.975	0.953	0.892	0.935	1.053	0.839	0.795	0.811
$h = 8$	0.953	0.971	0.945	0.741	0.972	1.106	0.779	0.762	0.763
$h = 12$	0.942	1.000	0.955	0.689	1.030	1.231	0.777	0.855	0.793

Japan	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.997	0.966	0.986	1.356	1.002	1.243	0.923	1.015	0.960
$h = 4$	1.025	0.924	0.943	0.753	0.971	0.841	0.852	0.900	0.850
$h = 8$	1.030	0.906	0.903	0.622	0.944	0.893	0.816	0.912	0.818
$h = 12$	1.018	0.909	0.891	0.789	0.954	1.280	0.818	0.979	0.847

UK	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.976	0.964	0.984	1.171	1.182	1.598*	0.951	1.146	0.986
$h = 4$	1.001	1.015	1.024	1.110	1.218	1.972*	0.994	1.048	1.000
$h = 8$	1.014	1.061	1.078	1.101	1.269*	2.661*	1.011	1.088	1.028
$h = 12$	1.017	1.061*	1.111	1.171	1.277*	3.521*	0.990	1.124	1.007

USA	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.003	1.091	1.032	1.257*	1.265	1.311	1.073	1.209*	1.123
$h = 4$	0.981	1.111	1.061*	1.089	1.154	1.017	1.024	1.045	1.053
$h = 8$	0.971	1.108	1.056*	0.995	1.105	0.876	0.992	0.975	0.985
$h = 12$	0.964	1.065	1.035*	1.036	1.034	0.856	0.947	0.902	0.915

CRPS

Canada	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.010	1.013	1.041	1.027	1.056	1.507*	1.013	1.075	1.042
$h = 4$	0.976	0.995	1.016	0.937	1.081	1.389*	0.953	0.979	0.986
$h = 8$	0.952	1.015	1.053	0.822	1.132	1.455*	0.877*	0.915	0.906*
$h = 12$	0.939*	1.004	1.084	0.804	1.100	1.584*	0.814*	0.873	0.844*

Germany	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.020	0.948	0.943	1.123	0.959	1.078	0.925	0.976	0.923
$h = 4$	1.063*	0.863*	0.817*	1.032	0.839*	0.882	0.770*	0.739*	0.762*
$h = 8$	1.066*	0.796*	0.761*	1.002	0.809*	0.877	0.690*	0.705	0.711*
$h = 12$	1.050*	0.761*	0.747*	1.053	0.863	1.008	0.677*	0.747*	0.721*

France	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.993	0.950	0.921*	1.026	0.911*	1.410*	0.908	0.957	0.887
$h = 4$	0.990	0.923	0.854*	0.853	0.834*	1.165	0.829*	0.847*	0.813*
$h = 8$	0.983	0.930	0.858*	0.752	0.873*	1.205	0.771*	0.784*	0.757*
$h = 12$	0.984	0.983	0.897	0.829	0.979	1.434*	0.775*	0.860*	0.795*

Italy	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.023	0.958	0.965	1.035	0.934	1.135	0.934	1.053	0.948
$h = 4$	1.021	0.983	0.962	0.901	0.928	1.006	0.838	0.782	0.803
$h = 8$	0.975	0.981	0.949	0.737	0.974	1.010	0.785	0.754	0.761
$h = 12$	0.954	1.030	0.979	0.675	1.051	1.098	0.774	0.833	0.787

Japan	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.987	0.998	0.980	1.624*	1.077	1.477*	0.980	1.146	1.041
$h = 4$	1.016	0.959	0.958	1.087	1.005	1.054	0.912	1.022	0.939
$h = 8$	1.011	0.938	0.933	0.926	0.972	1.084	0.867	0.994	0.891
$h = 12$	0.973	0.910	0.905	1.065	0.957	1.323	0.831	0.994	0.868

UK	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.984	1.005	1.004	1.405*	1.255*	1.884*	0.987	1.228*	1.043
$h = 4$	1.002	1.038	1.055*	1.175	1.260	1.789*	1.026	1.134	1.055
$h = 8$	1.017	1.080	1.108*	1.146	1.360*	2.073*	1.036	1.144	1.070
$h = 12$	1.021	1.085*	1.140*	1.150	1.386*	2.321*	1.019	1.172	1.054

USA	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.996	1.118	1.040	1.383*	1.313*	1.358*	1.069	1.209*	1.115
$h = 4$	0.968	1.167	1.071*	1.132	1.207	1.054	1.047	1.053	1.074
$h = 8$	0.960	1.189	1.067	0.994	1.174	0.961	1.041	1.008	1.034
$h = 12$	0.946	1.127	1.044	0.982	1.076	0.999	0.981	0.910	0.944

Notes: The models are detailed in Table 2. For each specification, the upper panel presents the ratios of RMSFEs relative to the CVAR benchmark. The lower panel presents the ratios of CRPSs relative to the CVAR benchmark. Values below 1 indicate the model outperforms the benchmark and vice versa. Gray shading indicates the best performing model. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano *t*-statistic with fixed-smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero are denoted by an asterisk, corresponding to the 5 percent significance level.

6.3. Investigating forecast performance over time

We have so far conducted a comprehensive evaluation of how different model specifications and prior choices affect forecast accuracy in the multi-country context, finding that multi-country VARs with hierarchical priors, in particular, the HS prior, are very helpful in forecasting inflation, as well as output growth and the interest rate. To get a better understanding of the source of the gains, we evaluate the models' forecasting performance over time. We plot in Figures 1-6 the cumulative sums of both RMSFEs and CRPSs at the selected horizons of 1,4,8, and 12 periods over the evaluation sample, averaged (arithmetic mean) across countries. Different colors with corresponding markers indicate different model specifications. The most recent Great Recession-financial crisis period (2007Q1-2009Q4) is highlighted in gray. For illustration, we only report results obtained from the benchmark and multi-country VARs with the three different scale mixtures of Normals priors (HS, NG, and NGG).

We first examine results obtained for output growth (Figures 1-2). The Great Recession-financial crisis clearly has a large effect on RMSFE and CRPS accuracy; all models' cumulative RMSFEs and CRPSs markedly increase after 2008. Before the crisis, the single-country CVAR benchmark performs similarly to multi-country VARs with hierarchical shrinkage and even does slightly better at long horizons. However, hierarchical shrinkage applied to multi-country VARs tends to be more beneficial after the crisis, which is particularly evident in the 4- and 8-steps-ahead density forecasts. Overall, in these aggregated measures, the NG specification is the best at short horizons, but the HS specification is better at long horizons, for both point and density forecasts.

When we compare the performance for inflation forecasts (Figures 3-4), there are several differences. First, compared to the results for output growth, we do not see as sharp an increase in cumulative RMSFEs and CRPSs during and after the financial crisis. There is some increase, but not as large as in the case of output growth forecasts. Second, when averaged across countries, before the crisis these models' forecasting performance is very similar at the 1-step-ahead horizon; after the crisis, the NG prior specification is slightly less accurate than the others. However, as the forecast horizon increases, the multi-country VAR with the HS prior becomes relatively more accurate, in both point and density forecasts. The light-tailed NG prior is clearly the worst among the three different scale mixtures of Normals priors and even worse than the single-country CVAR benchmark, particularly at longer horizons.

Moving to interest rate forecasts (Figures 5-6), the effects of the Great Recession-financial crisis are clear at multi-step forecast horizons, but less dramatic than for output growth forecasts. As interest rates in all G7 countries hit their effective lower bound, all models have difficulties in capturing the abrupt changes in short-term interest rates. At the 1-step-ahead horizon, the performance of the models is broadly similar, with the exception of the light-tailed NG prior; although comparable to others, the HS specification has slightly better accuracy. As the forecast horizon increases, the benefits obtained from the multi-country VARs with hierarchical priors become more evident; these specifications clearly outperform the single-country CVAR benchmark. The HS prior is better than the more flexible NGG prior, while the NG prior is clearly the worst among the three scale mixtures of Normals priors.

To conclude, we confirm that hierarchical shrinkage in multi-country VARs, especially coupled with the HS prior, delivers more accurate and robust forecasts over time for all three target variables. In these results aggregated across countries, for output growth, the NG prior is more preferable at short horizons, but the HS prior does better at long horizons. Gains are mainly obtained in the post-crisis evaluation period. For inflation, gains from the multi-country VAR with the HS prior are more evident as the forecast horizon increases. For the short-term interest rate, all models show difficulties in obtaining accurate forecasts as the horizon increases. Hierarchical shrinkage in multi-country VARs is generally better than the single-country benchmark as the horizon increases, and the HS prior tends to be more beneficial than the other scale mixtures of Normals priors.

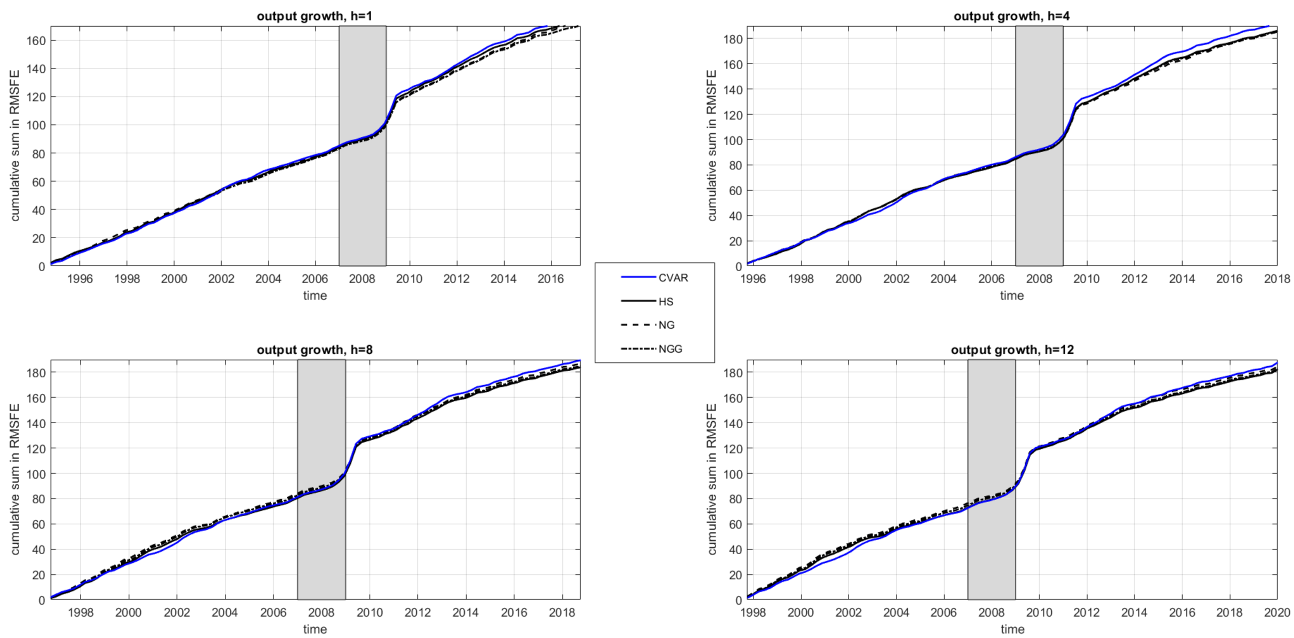

Figure 1: The figure presents cumulative sums (taken over time and averaged across countries) of RMSFEs for output growth forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 2.

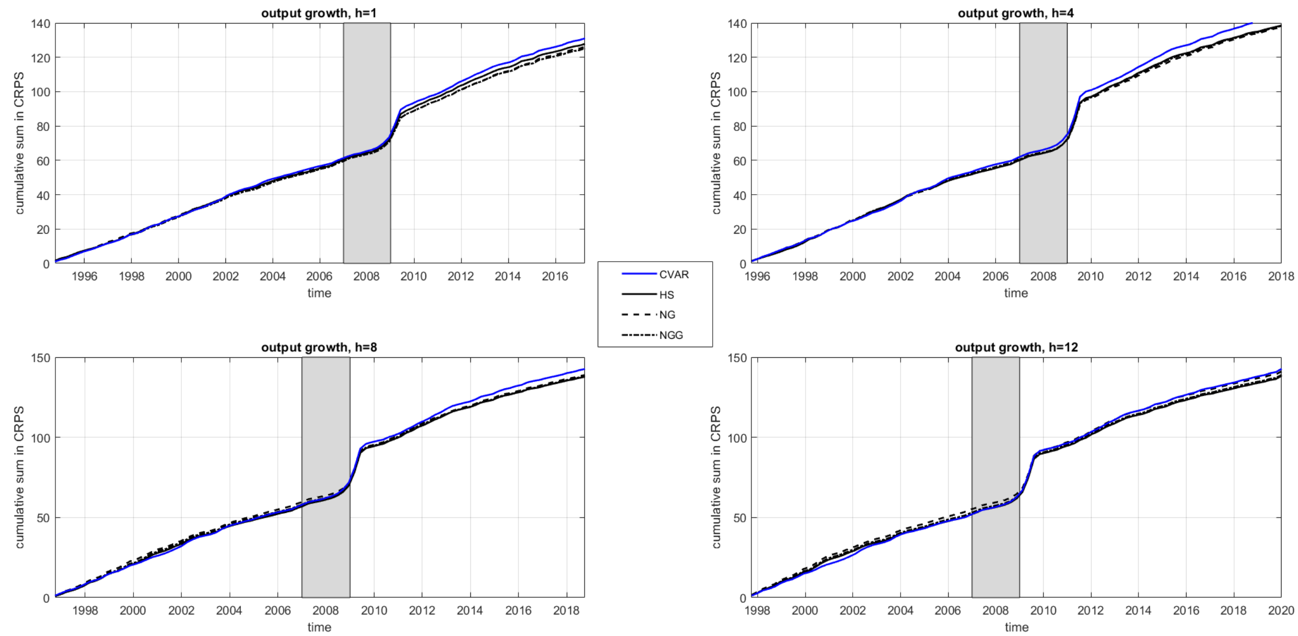

Figure 2: The figure presents cumulative sums (taken over time and averaged across countries) of CRPSs for output growth forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 2.

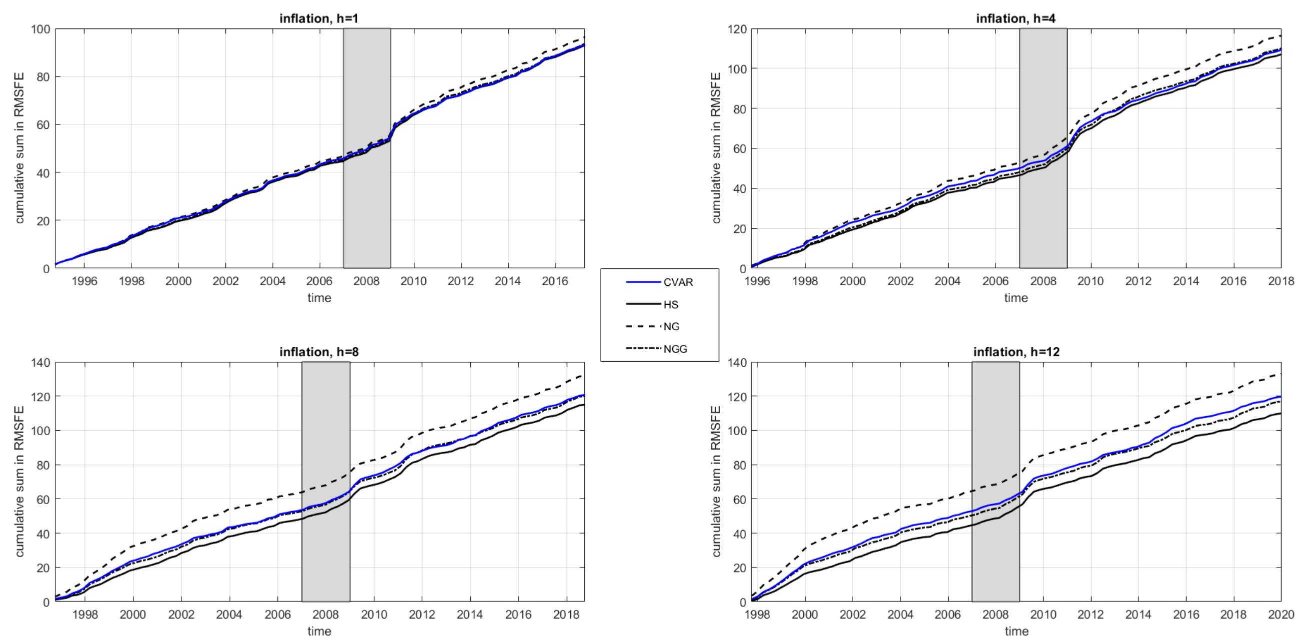

Figure 3: The figure presents cumulative sums (taken over time and averaged across countries) of RMSFEs for inflation forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 2.

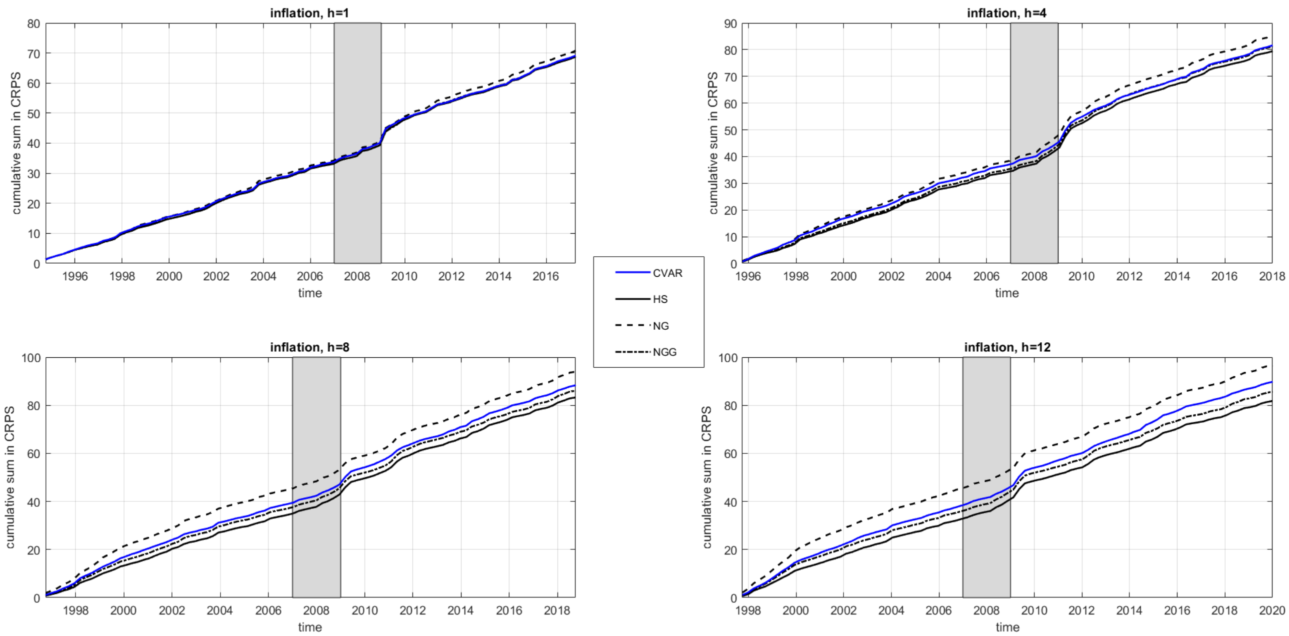

Figure 4: The figure presents cumulative sums (taken over time and averaged across countries) of CRPSs for inflation forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 2.

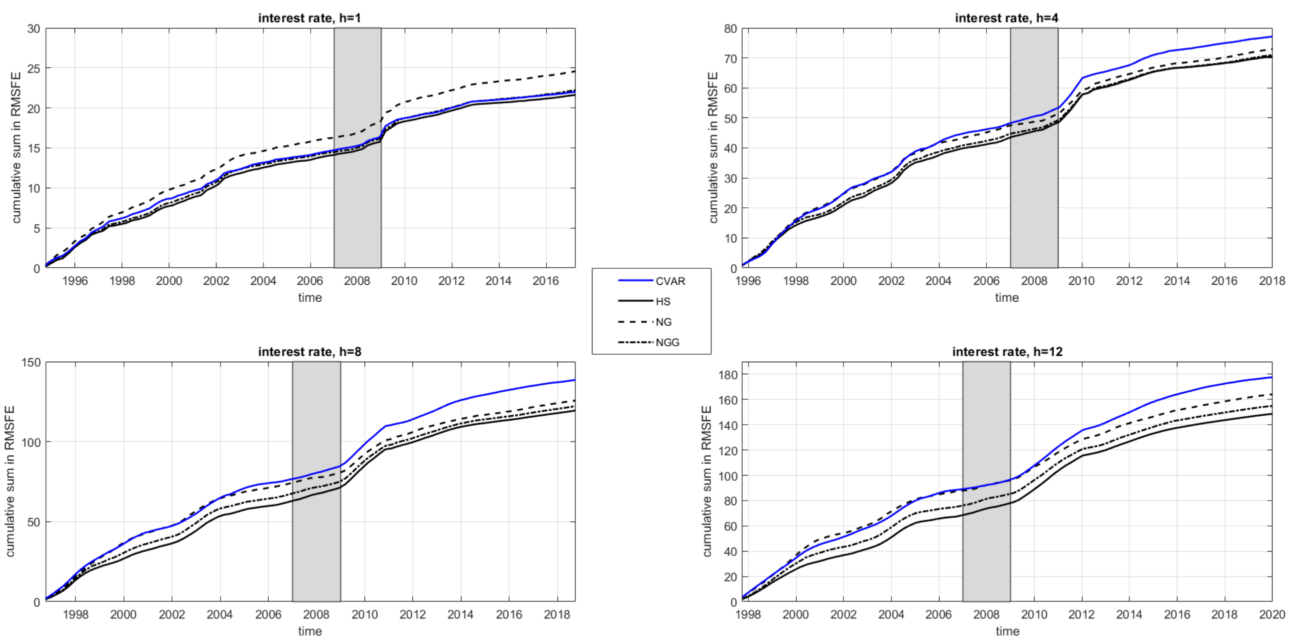

Figure 5: The figure presents cumulative sums (taken over time and averaged across countries) of RMSFEs for interest rate forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 2.

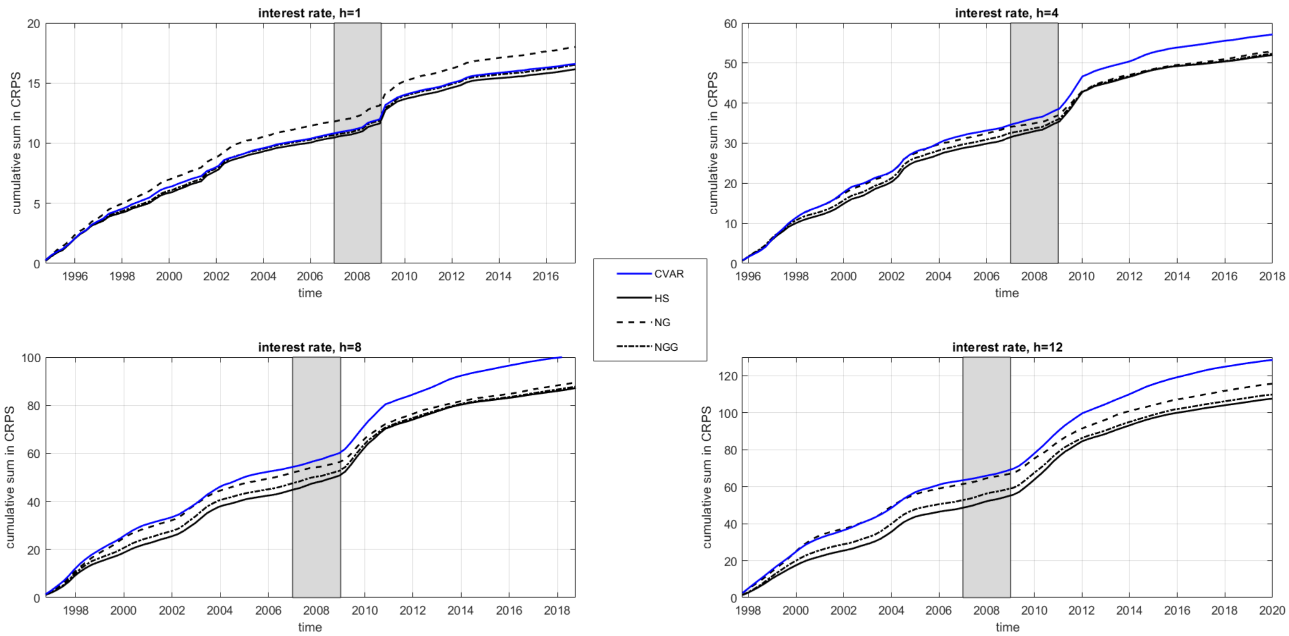

Figure 6: The figure presents cumulative sums (taken over time and averaged across countries) of CRPSs for interest rate forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 2.

6.4. Some robustness checks

In this subsection, we conduct several robustness checks of our main results presented above. We focus on the multi-country VAR with the Horseshoe prior (the HS specification), since overall it delivers the best forecast performance. In the interest of space, we briefly summarize the main findings; the details of the results can be found in Appendix D.

Prior grouping of coefficients. As a check of the baseline prior's grouping of coefficients, we consider three alternative groupings of coefficients compared to the one used in the main results. First, we group all coefficients related to CSH together (HS-CSH) and assume that the priors of elements in vec(β^{CSH}) follow (9). Second, we do not make any attempt to search for restrictions but instead group both the intercepts and all autoregressive coefficients together (HS-A) and assume that all coefficients follow the same prior distributions as in (9). This has been examined in the single-country Bayesian VAR context by Cross, Hou, and Poon (2020), who find that scale mixtures of Normals priors do not improve the forecasting accuracy compared to conventional Minnesota priors. Finally, we group coefficients based on equations (HS-E) by assuming that priors for coefficients in each equation of (1) are the same. The equation-based shrinkage priors are more often seen in single-country Bayesian VARs. Follett and Yu (2019) and Huber and Feldkircher (2019) use this type of specification for the HS and NG priors, respectively. Cadonna, Frühwirth-Schnatter, and Knaus (2020) also propose an equation-wise specification for single-country time-varying parameter VARs.

In Tables D.1-D.3, we report summary statistics on the percentage of gains for the alternative specifications described above compared to the specifications we use in the main results for all horizons, short horizons ($h \leqslant 6$), and long horizons ($h > 6$). The results can be summarized as follows. First, the HS-CSH, HS-A, and HS-E priors all improve forecast accuracy for the interest rate, especially at long horizons, with average gains of more than 10 percent. Even though neither HS-A nor HS-E considers the underlying structure of model parameters, for the interest rate they forecast better than the HS-CSH specification. Second, these three alternatives are not helpful in forecasting output growth. The average gains are all negative, and they lead to loss of forecast accuracy in more than 75 percent of all cases. The HS-A prior has the overall worst forecast performance for output growth, and while output growth forecast performance from HS-CSH and HS-E is roughly similar, these approaches are outperformed by the HS specification used in the main results. Third, for inflation, the HS-CSH prior delivers forecast performance similar to our HS specification. The average gains are close to zero; in around half of all cases, the gains are positive. However, the HS-A and HS-E specifications generally reduce forecast accuracy (with HS-A the worst alternative), with impacts that are negative in more than 70 percent of all cases, and average losses of roughly 2 percent. Overall, these results suggest that it is not possible to improve our overall results by modifying our baseline grouping of coefficients.

Stochastic volatility. As another check, we also assess whether stochastic volatility is useful to improve forecast accuracy in the multi-country context. We modify the distributional assumption of u_t in (2) by assuming that $u_t \sim i.i.d.\ N(0, \Sigma)$. We specify a Normal prior for the VAR's coefficients and an Inverse Wishart prior for Σ and use the corrected triangular algorithms proposed in Carriero, et al. (2022) to estimate and forecast. In Table D.4, we provide summary statistics on the percentage differences in the accuracy of forecasts from the models with and without SV. The results clearly indicate the usefulness of stochastic volatility. The constant volatility models are outperformed by stochastic volatility models for all horizons and all target variables. The benefits are particularly evident in forecasting inflation and the interest rate. For inflation, introducing stochastic volatility in multi-country VARs delivers gains in all cases at long horizons, and average gains are large: 34 percent for point forecasts and 27 percent for density forecasts. For the interest rate, stochastic volatility is more beneficial at short horizons. The average gains are around 12 (20) percent for point (density) forecasts, and gains are positive in nearly 80 (85) percent of cases for point (density) forecasts.

Estimation scheme. There is a long debate on the relative forecast performance of rolling and expanding window (recursive scheme) estimation in the forecasting literature. While rolling window estimates can be more robust to structural breaks, expanding window parameter estimates can be more efficient, helping forecast precision. In Table D.5, we compare point and density forecasts from rolling and recursive schemes, taking as a benchmark the recursive scheme used in the paper's main results. The rolling scheme results use a window of 22 years of data, corresponding to the size of the sample used to generate the first forecast observation in our main results. On average, the rolling scheme forecasts are slightly better than the recursive forecasts, but the two methods perform broadly similarly. Compared to the recursive baseline, average gains to the rolling scheme are small and generally not statistically significant. By looking at the percentage of cases in which a given method outperforms the other, it appears that the rolling scheme does relatively better for inflation, for point forecasts for output growth, and interest rate forecasts at long horizons.

Univariate forecasting benchmarks. To understand the relative merits of our models with respect to univariate models, which are often tough benchmarks in the forecasting literature, we compare forecasts from the HS specification to those from univariate models with SV. We choose AR(p) models for output growth and the interest rate, with p = 2 and 4 lags, respectively, following Clark and Ravazzolo (2015). For inflation, we choose an unobserved component model with SV as in Chan (2018). Table D.6 provides summary statistics for these accuracy comparisons. In these results, the multi-country HS specification yields more accurate forecasts of inflation and the interest rate. The average gains for the interest rate exceed 7 percent. Gains are positive in more than 90 percent of cases for density forecasts of inflation. For output growth, average gains from the HS specification are small but still positive. Overall, our main results based on a single-country VAR baseline still obtain when the baseline is changed to common univariate models.

Effective lower bound on interest rates. Since the 2007-2009 financial crisis put interest rates in all G7 countries at the effective lower bound for a number of years, one concern is that our reduced-form VAR models may forecast interest rates to be much higher than actual rates.[11] In Figures D.1 and D.2, we present point forecasts and associated 95 percent interval forecasts of the interest rate for all G7 countries obtained from the multi-country VAR with the Horseshoe prior (the HS specification) at horizons of 1 and 12 steps, respectively. We find that the ELB does not seem to be a major concern for short horizon (1-step-ahead) forecasts, as our model is able to track the true interest rate rather well even during the ELB period. However, some bias in the forecasts emerges during the ELB period when we look at 12-steps-ahead forecasts. Our model generally predicts the interest rate to be higher than the actual rate. While more evident in the ELB period, the problem is present even before the ELB period. However, in the case of Japan, we see that our model forecasts the interest rate to be much higher than the true value early in the sample, but the forecasts gradually decline and are able to track the realized values fairly well for much of the sample. We conclude from these results that the ELB has some impact on our interest rate forecasting results — as it likely will for most any VAR — but does not necessarily entirely distort them.

Directional forecasts. To provide a rough gauge on whether our forecasting models are also able to predict turning points, we use the nonparametric test developed by Pesaran and Timmermann (1992) to assess directional forecast performance for (1-step-ahead) changes in output growth. In Table D.7, we report test statistics and associated p-values for all 7 countries from both the multi-country VAR-SV

[11]Japan hit the ELB earlier than other countries. The short-term interest rate in Japan remains around zero in the entire forecast evaluation period.

model with the Horseshoe prior and the single-country VAR-SV benchmark. The results show that, except for Canada, for the multi-country VAR-SV with Horseshoe prior, the test strongly rejects the null of no predictability. However, for the single-country VAR-SV benchmark, we cannot reject the null for all 7 countries. This provides clear evidence that the multi-country VAR-SV model with Horseshoe prior also has better predictive power for the directional forecasts of output growth.

Alternative hierarchical shrinkage in country-specific VAR-SV. In the previous subsections, we used the Minnesota-type adaptive hierarchical prior in Chan (2021), a published prior specification, as a competitor to the multi-country VAR-SV with hierarchical shrinkage. However, when comparing it to the results presented in previous subsections, there are some differences in terms of both specification and estimation. First, Chan (2021) uses the Normal-Gamma prior, but our preferable prior choice is the Horseshoe prior. Second, Chan (2021) introduces additional pre-specified hyperparameters as in the standard Minnesota-type prior, but in our specification all local scale parameters are learned from the data. Finally, Chan (2021) estimates the model in structural form, but we estimate the models in reduced form.

As a robustness check, we estimate country-specific VAR-SV models in reduced form with hierarchical shrinkage. We use the Horseshoe prior as specified in (10) with three groups of coefficients as in the standard Minnesota-type prior: intercepts, coefficients related to own lags, and cross variable lags. We compare forecasts with CVAR-H as in Chan (2021) and HS as we use in the main results. Tables D.8 and D.9 provide summary statistics for these accuracy comparisons. Clearly, our hierarchical prior specification provides more accurate forecasts compared to Chan (2021). The average gains are large (more than 7.5 percent) for interest rate forecasts. Gains are positive in at least more than 75 percent of cases. When comparing forecasts with multi-country VAR-SV combined with HS, we find that the multi-country VAR-SV achieves positive average gains in nearly all cases (except point forecasts for interest rates at short horizons). Gains are more evident for density forecasts of inflation and interest rate forecasts at long horizons. These results further confirm the benefits of adding foreign information to country-specific VAR-SV models.

7. Conclusions

In this paper, we use hierarchical shrinkage in multi-country Bayesian VARs and examine its macroeconomic forecasting ability. In implementation, we consider three different scale mixtures of Normals priors, namely the Horseshoe prior, the Normal-Gamma prior, and the Normal-Gamma-Gamma prior, which have been shown to benefit macroeconomic forecasting in single-country settings. We also provide some new theoretical results for the Normal-Gamma prior. Empirically, we compare the forecast accuracy with country-specific VARs, country-specific factor-augmented VARs, global VARs, and alternative shrinkage approaches for multi-country VARs that have been used in the macroeconomic forecasting literature. All of our models include stochastic volatility, which is helpful to forecast accuracy. We confirm the benefits of enlarging single-country information sets to include information across countries. Hierarchical shrinkage in the multi-country VAR model with the Horseshoe prior has the overall best forecast performance.

References

Abramowitz, Milton, and Irene A. Stegun (1965), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Washington: US Government Printing Office.

Angelini, Elena, Magdalena Lalik, Michele Lenza, and Joan Paredes (2019), "Mind the gap: a multi-country BVAR benchmark for the eurosystem projections," *International Journal of Forecasting*, 35, 1658–1668, https://doi.org/10.1016/j.ijforecast.2018.12.004.

Armagan, Artin, Merlise Clyde, and David B. Dunson (2011), "Generalized beta mixtures of Gaussians," in *Advances in Neural Information Processing Systems*, 523–531.

Bai, Jushan, and Serena Ng (2002), "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221, https://doi.org/10.1111/1468-0262.00273.

Bitto, Angela, and Sylvia Frühwirth-Schnatter (2019), "Achieving shrinkage in a time-varying parameter model framework," *Journal of Econometrics*, 210, 75–97, https://doi.org/10.1016/j.jeconom.2018.11.006.

Brooks, Stephen P., and Andrew Gelman (1998), "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455, https://doi.org/10.1080/10618600.1998.10474787.

Cadonna, Annalisa, Sylvia Frühwirth-Schnatter, and Peter Knaus (2020), "Triple the Gamma — a unifying shrinkage prior for variance and variable selection in sparse state space and TVP models," *Econometrics*, 8, 1–36, https://doi.org/10.3390/econometrics8020020.

Canova, Fabio, and Matteo Ciccarelli (2009), "Estimating multicountry VAR models," *International Economic Review*, 50, 929–959, https://doi.org/10.1111/j.1468-2354.2009.00554.x.

——— (2013), "Panel vector autoregressive models: A survey," in *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*, 32 of Advances in Econometrics: Emerald Group Publishing Limited, 205–246, https://doi.org/10.1108/S0731-9053(2013)0000031006.

Canova, Fabio, Matteo Ciccarelli, and Eva Ortega (2007), "Similarities and convergence in G-7 cycles," *Journal of Monetary Economics*, 54, 850–878, https://doi.org/10.1016/j.jmoneco.2005.10.022.

Carriero, Andrea, Joshua Chan, Todd E. Clark, and Massimiliano Marcellino (2022), "Corrigendum to: Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, forthcoming, https://doi.org/10.1016/j.jeconom.2021.11.010.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019), "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154, https://doi.org/10.1016/j.jeconom.2019.04.024.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2009), "Handling sparsity via the horseshoe," in *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, 73–80: PMLR, April, https://proceedings.mlr.press/v5/carvalho09a.html.

——— (2010), "The horseshoe estimator for sparse signals," *Biometrika*, 97, 465–480, https://doi.org/10.1093/biomet/asq017.

Chan, Joshua C.C. (2017), "Notes on Bayesian macroeconometrics," http://joshuachan.org/papers/BayesMacro.pdf.

——— (2018), "Specification tests for time-varying parameter models with stochastic volatility," *Econometric Reviews*, 37, 807–823, https://doi.org/10.1080/07474938.2016.1167948.

——— (2021), "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37, 1212–1226, https://doi.org/10.1016/j.ijforecast.2021.01.002.

Clark, Todd E., and Francesco Ravazzolo (2015), "Macroeconomic forecasting performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30, 551–575, https://doi.org/10.1002/jae.2379.

Cogley, Timothy, and Thomas J. Sargent (2005), "Drifts and volatilities: Monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302, https://doi.org/10.1016/j.red.2004.10.009.

Coroneo, Laura, and Fabrizio Iacone (2020), "Comparing predictive accuracy in small samples using fixed-smoothing asymptotics," *Journal of Applied Econometrics*, 35, 391–409, https://doi.org/10.1002/jae.2756.

Cross, Jamie L., Chenghan Hou, and Aubrey Poon (2020), "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *International Journal of Forecasting*, 36, 899–915, https://doi.org/10.1016/j.ijforecast.2019.10.002.

Cuaresma, Jesús Crespo, Martin Feldkircher, and Florian Huber (2016), “Forecasting with global vector autoregressive models: A Bayesian approach,” *Journal of Applied Econometrics*, 31, 1371–1391, https://doi.org/10.1002/jae.2504.

D’Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), “Macroeconomic forecasting and structural change,” *Journal of Applied Econometrics*, 28, 82–101, https://doi.org/10.1002/jae.1257.

Dées, Stéphane, and Jochen Güntner (2017), “Forecasting inflation across euro area countries and sectors: A panel VAR approach,” *Journal of Forecasting*, 36, 431–453, https://doi.org/10.1002/for.2444.

Del Negro, Marco, and Giorgio E. Primiceri (2015), “Time varying structural vector autoregressions and monetary policy: A corrigendum,” *Review of Economic Studies*, 82, 1342–1345, https://doi.org/10.1093/restud/rdv024.

Diebold, Francis X., and Robert S. Mariano (1995), “Comparing predictive accuracy,” *Journal of Business and Economic Statistics*, 13, 253–263, https://doi.org/10.2307/1392185.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984), “Forecasting and conditional projection using realistic prior distributions,” *Econometric Reviews*, 3, 1–100, https://doi.org/10.1080/07474938408800053.

Dovern, Jonas, Martin Feldkircher, and Florian Huber (2016), “Does joint modelling of the world economy pay off? Evaluating global forecasts from a Bayesian GVAR,” *Journal of Economic Dynamics and Control*, 70, 86–100, https://doi.org/10.1016/j.jedc.2016.06.006.

Feldkircher, Martin, Florian Huber, Gary Koop, and Michael Pfarrhofer (2021), “Approximate Bayesian inference and forecasting in huge-dimensional multi-country VARs,”Technical Report 2103.04944, arXiv.org, https://ideas.repec.org/p/arx/papers/2103.04944.html.

Follett, Lendie, and Cindy Yu (2019), “Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior,” *Econometrics and Statistics*, 11, 130–144, https://doi.org/10.1016/j.ecosta.2018.12.004.

Gefang, Deborah, Gary Koop, and Aubrey Poon (2022), “Variational Bayesian inference in large vector autoregressions with hierarchical shrinkage,” *International Journal of Forecasting*, forthcoming, https://doi.org/10.1016/j.ijforecast.2021.11.012.

George, Edward I., Dongchu Sun, and Shawn Ni (2008), “Bayesian stochastic search for VAR model restrictions,” *Journal of Econometrics*, 142, 553–580, https://doi.org/10.1016/j.jeconom.2007.08.017.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2015), "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97, 436–451, https://doi.org/10.1162/REST_a_00483.

——— (2021), "Economic predictions with big data: The illusion of sparsity," *Econometrica*, 89, 2409–2437, https://doi.org/10.3982/ECTA17842.

Giannone, Domenico, and Lucrezia Reichlin (2009), "Comments on 'Forecasting economic and financial variables with global VARs'," *International Journal of Forecasting*, 25, 684–686, https://doi.org/10.1016/j.ijforecast.2009.06.003.

Gneiting, Tilmann, and Adrian E. Raftery (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359–378, https://doi.org/10.1198/016214506000001437.

Griffin, Jim, and Phil Brown (2017), "Hierarchical shrinkage priors for regression models," *Bayesian Analysis*, 12, 135–159, https://doi.org/10.1214/15-BA990.

Griffin, Jim, and Philip Brown (2010), "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5, 171–188, https://doi.org/10.1214/10-BA507.

Huber, Florian (2016), "Density forecasting using Bayesian global vector autoregressions with stochastic volatility," *International Journal of Forecasting*, 32, 818–837, https://doi.org/10.1016/j.ijforecast.2015.12.008.

Huber, Florian, and Martin Feldkircher (2019), "Adaptive shrinkage in Bayesian vector autoregressive models," *Journal of Business and Economic Statistics*, 37, 27–39, https://doi.org/10.1080/07350015.2016.1256217.

Kapetanios, George, Massimiliano Marcellino, and Fabrizio Venditti (2019), "Large time-varying parameter VARs: a non-parametric approach," *Journal of Applied Econometrics*, 34, 1027–1049, https://doi.org/10.1002/jae.2722.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65, 361–393, https://doi.org/10.1111/1467-937X.00050.

Koop, Gary, and Dimitris Korobilis (2016), "Model uncertainty in panel vector autoregressive models," *European Economic Review*, 81, 115–131, https://doi.org/10.1016/j.euroecorev.2015.09.006.

——— (2019), "Forecasting with high-dimensional panel VARs," *Oxford Bulletin of Economics and Statistics*, 81, 937–959, https://doi.org/10.1111/obes.12303.

Koop, Gary, Roberto Leon-Gonzalez, and Rodney W. Strachan (2009), "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control*, 33, 997–1017, https://doi.org/10.1016/j.jedc.2008.11.003.

Koop, Gary, Stuart McIntyre, James Mitchell, and Aubrey Poon (2020), "Regional output growth in the United Kingdom: More timely and higher frequency estimates from 1970," *Journal of Applied Econometrics*, 35, 176–197, https://doi.org/10.1002/jae.2748.

Korobilis, Dimitris (2016), "Prior selection for panel vector autoregressions," *Computational Statistics and Data Analysis*, 101, 110–120, https://doi.org/10.1016/j.csda.2016.02.011.

Korobilis, Dimitris, and Davide Pettenuzzo (2019), "Adaptive hierarchical priors for high-dimensional vector autoregressions," *Journal of Econometrics*, 212, 241–271, https://doi.org/10.1016/j.jeconom.2019.04.029.

Litterman, Robert B. (1986), "Forecasting with Bayesian vector autoregressions—five years of experience," *Journal of Business and Economic Statistics*, 4, 25–38, https://doi.org/10.1080/07350015.1986.10509491.

Makalic, Enes, and Daniel F. Schmidt (2015), "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23, 179–182, https://doi.org/10.1109/LSP.2015.2503725.

Mohaddes, Kamiar, and Mehdi Raissi (2020), "Compilation, revision, and updating of the global VAR (GVAR) database, 1979q2-2019q4," https://doi.org/10.17863/CAM.56762.

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima (2007), "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425–449, https://doi.org/10.1016/j.jeconom.2006.07.008.

Park, Trevor, and George Casella (2008), "The Bayesian Lasso," *Journal of the American Statistical Association*, 103, 681–686, https://doi.org/10.1198/016214508000000337.

Pesaran, M. Hashem, Til Schuermann, and L. Vanessa Smith (2009), "Forecasting economic and financial variables with global VARs," *International Journal of Forecasting*, 25, 642–675, https://doi.org/10.1016/j.ijforecast.2009.08.007.

Pesaran, M. Hashem, and Allan Timmermann (1992), "A simple nonparametric test of predictive performance," *Journal of Business and Economic Statistics*, 10, 461–465, https://doi.org/10.1080/07350015.1992.10509922.

Primiceri, Giorgio E. (2005), “Time varying structural vector autoregressions and monetary policy,” *Review of Economic Studies*, 72, 821–852, https://doi.org/10.1111/j.1467-937X.2005.00353.x.

Roberts, Gareth O., and Jeffrey S. Rosenthal (2009), “Examples of adaptive MCMC,” *Journal of Computational and Graphical Statistics*, 18, 349–367, https://doi.org/10.1198/jcgs.2009.06134.

Smith, Michael S., and Shaun P. Vahey (2016), “Asymmetric forecast densities for US macroeconomic variables from a Gaussian copula model of cross-sectional and serial dependence,” *Journal of Business and Economic Statistics*, 34, 416–434, https://doi.org/10.1080/07350015.2015.1044533.

Stock, James H., and Mark W. Watson (2002a), “Forecasting using principal components from a large number of predictors,” *Journal of the American Statistical Association*, 97, 1167–1179, https://doi.org/10.1198/016214502388618960.

——— (2002b), “Macroeconomic forecasting using diffusion indexes,” *Journal of Business and Economic Statistics*, 20, 147–162, https://doi.org/10.1198/073500102317351921.

Appendix A. Proof of Theorem 1

In the following proof, the notation $\sim$ indicates asymptotic equivalence. We say that a is asymptotically equivalent to b if $a/b = O(1)$.

As shown in equation (14) of Bitto and Frühwirth-Schnatter (2019), the marginal density for $\beta_j \sim \mathcal{NG}(\lambda, \kappa)$, given λ, κ, can be expressed as

$$\pi_{NG}(\beta_j) = \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right),$$

where $K_p(\cdot)$ is the modified Bessel function of the second kind of index p. Let us first consider the concentration properties. If $\lambda > \frac{1}{2}$, then $\lambda - \frac{1}{2} > 0$. By 9.6.9 in Abramowitz and Stegun (1965), as $\left|\beta_j\right| \to 0$,

$$K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right) \sim \frac{1}{2}\Gamma(\lambda - \frac{1}{2})\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda}.$$

Then,

$$\begin{aligned}\pi_{NG}(\beta_j) &\sim \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}} \times \frac{1}{2}\Gamma(\lambda - \frac{1}{2})\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda} \\ &= \frac{\sqrt{\lambda\kappa}}{\sqrt{\pi}}\frac{\Gamma(\lambda - \frac{1}{2})}{\Gamma(\lambda)} \times \frac{1}{2} = O(1).\end{aligned}$$

If $0 < \lambda < \frac{1}{2}$, recall that

$$K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right) = \frac{1}{2}\pi\frac{I_{\frac{1}{2}-\lambda}(\sqrt{\lambda\kappa}\left|\beta_j\right|) - I_{\lambda-\frac{1}{2}}(\sqrt{\lambda\kappa}\left|\beta_j\right|)}{\sin\left((\lambda - \frac{1}{2})\pi\right)},$$

where $I_p(\cdot)$ is the modified Bessel function of the first kind with index p. By 9.6.7 in Abramowitz and Stegun (1965), as $\left|\beta_j\right| \to 0$,

$$\begin{aligned}I_{\frac{1}{2}-\lambda}(\sqrt{\lambda\kappa}\left|\beta_j\right|) &\sim \frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda}}{\Gamma(\frac{3}{2}-\lambda)} \\ I_{\lambda-\frac{1}{2}}(\sqrt{\lambda\kappa}\left|\beta_j\right|) &\sim \frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\lambda-\frac{1}{2}}}{\Gamma(\frac{3}{2}-\lambda)}.\end{aligned}$$

Thus,

$$\pi_{NG}(\beta_j) \sim \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}} \times \frac{1}{2}\pi \times \frac{1}{\sin((\lambda-\frac{1}{2})\pi)} \times \left(\frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda}}{\Gamma(\frac{3}{2}-\lambda)} - \frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\lambda-\frac{1}{2}}}{\Gamma(\frac{3}{2}-\lambda)}\right)$$

$$= C - \frac{(\sqrt{\lambda\kappa})^{2\lambda} \times (\frac{1}{2})^{2\lambda}\sqrt{\pi}}{\sin((\lambda-\frac{1}{2})\pi)\Gamma(\lambda)\Gamma(\lambda+\frac{1}{2})}\left(\frac{1}{\left|\beta_j\right|}\right)^{1-2\lambda}$$

$$= O\left(\left(\frac{1}{\left|\beta_j\right|}\right)^{1-2\lambda}\right).$$

Finally, if $\lambda = \frac{1}{2}$, by 9.6.8 in Abramowitz and Stegun (1965),

$$K_0\left(\sqrt{\frac{1}{2}\kappa}\left|\beta_j\right|\right) \sim -\log(\sqrt{\frac{1}{2}\kappa}\left|\beta_j\right|).$$

Then,

$$\pi_{NG}(\beta_j) \sim \frac{\sqrt{\frac{1}{2}\kappa}}{\pi} \times -\log(\sqrt{\frac{1}{2}\kappa}\left|\beta_j\right|) = O\left(\frac{1}{\log(\left|\beta_j\right|)}\right).$$

We now move to the asymptotic tail behavior. By 9.7.2 in Abramowitz and Stegun (1965), as $\left|\beta_j\right| \to \infty$,

$$K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right) \sim \sqrt{\frac{\pi}{2\sqrt{\lambda\kappa}\left|\beta_j\right|}}e^{-\sqrt{\lambda\kappa}\left|\beta_j\right|}.$$

Then,

$$\pi_{NG}(\beta_j) \sim \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}}\sqrt{\frac{\pi}{2}}(\sqrt{\lambda\kappa})^{-\frac{1}{2}}\left|\beta_j\right|^{-\frac{1}{2}}\exp(-\sqrt{\lambda\kappa}\left|\beta_j\right|)$$

$$= O\left(\frac{\left|\beta_j\right|^{\lambda-1}}{\exp(\sqrt{\lambda\kappa}\left|\beta_j\right|)}\right),$$

which completes the proof.

Appendix B. Model specifications and priors

Appendix B.1. Country-specific VARs

The country-specific VAR(p) model — denoted the CVAR specification in the paper's results — is specified as

$$y_{i,t} = c_i + \sum_{l=1}^{p} B_{i,l} y_{i,t-l} + u_{i,t} \tag{B.1}$$

$$u_{i,t} = A_i^{-1} H_{i,t}^{0.5} \epsilon_{i,t}, \epsilon_{i,t} \sim i.i.d.\ N(0, I_G), \tag{B.2}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, and the dimension of $y_{i,t}$, $u_{i,t}$ and $\epsilon_{i,t}$ is $G \times 1$. A_i^{-1} is a lower triangular matrix with diagonal elements equal to 1. $H_{i,t}$ is diagonal with generic j-th element $h_{ij,t}$ evolving as a random walk (RW):

$$\ln h_{ij,t} = \ln h_{ij,t-1} + e_{ij,t}, j = 1, \ldots, G, \tag{B.3}$$

where $e_{it} \sim N(0, \Phi_i)$ with a full covariance matrix Φ_i as in Primiceri (2005).

Letting $B_i = [c_i, B_{i,1}, \ldots, B_{i,p}]'$, the priors are specified as:

$$vec(B_i) \sim N(0, \underline{\Omega}_{B_i})$$
$$vec(A_i) \sim N(0, \underline{\Omega}_{A_i})$$
$$\Phi_i \sim IW(Q_0, W_0).$$

For the prior variances of the autoregressive coefficient matrices, we set them as in the Minnesota prior:

$$\underline{\Omega}_{B_{i,l}^{(mn)}} = \begin{cases} \frac{\lambda_1}{l^{\lambda_3}} \frac{1}{\sigma_n^2} & \text{for the coefficients on own lags} \\ \frac{\lambda_2}{l^{\lambda_3}} \frac{\sigma_m^2}{\sigma_n^2} & \text{for the coefficients on cross-variable lags} \\ \lambda_0 \sigma_m^2 & \text{for the intercept,} \end{cases} \tag{B.4}$$

where $m, n = 1, \ldots, G$. λ_1 measures the overall tightness to coefficients related to own lags. λ_2 is related to cross-variable shrinkage. We assume Gamma priors for these two hyperparameters: $\lambda_1 \sim \mathcal{G}(1, 0.04)$, $\lambda_2 \sim \mathcal{G}(1, 0.04^2)$. λ_3 determines the additional shrinkage for coefficients associated with higher order lags and is set to 2 (quadratic decay). The scale parameters σ_m^2, σ_n^2 are obtained from univariate AR(1) regressions. We elicit an uninformative prior for the intercept by setting $\lambda_0 = 100$. In the case of the free elements in the contemporaneous matrix A_i, we set the prior mean to 0 and the prior variance to be non-informative: $\underline{\Omega}_{A_i} = 10 \times I$. Finally, as in the previous section, we follow the literature and set a modestly informative prior for Φ: $\Phi \sim IW(Q_0, W_0)$, where Q_0, W_0 take very conservative values: $W_0 = 0.01 \times I$ and

$Q_0 = G + 2$.[12]

For the country-specific VAR with hierarchical shrinkage (CVAR-H), we follow exactly the approach in Chan (2021). Following Chan, the reduced-form model (B.1) is expressed in structural form

$$A_i y_{i,t} = \tilde{c}_i + \sum_{l=1}^{p} \tilde{B}_{i,l} y_{i,t-l} + H_{i,t}^{0.5} \epsilon_{i,t},$$

where $\tilde{c}_i = Ac_i$, $\tilde{B}_{i,l} = AB_{i,l}$, and the innovations $e_{ij,t}$ in (B.3) are assumed to be independent across variables (equation $j = 1, \ldots, G$ of the VAR for country i): $e_{ij,t} \sim N(0, \sigma^2_{e_{ij}})$. The priors are specified as

$$\beta_{i,j} | \lambda_1, \lambda_2, \psi_{i,j}, C_{i,j} \sim N(0, 2\lambda_{i,j}\psi_{i,j}C_{i,j}),$$

where $\lambda_{i,j}$ equals λ_1 if $\beta_{i,j}$ are related to own lags but equals λ_2 for coefficients related to cross-variable lags. $C_{i,j}$ are specified according to

$$C_{i,l} = \left\{ \begin{array}{ll} \frac{1}{l^{\lambda_3}} \frac{1}{\sigma_n^2} & \text{for the coefficients on own lags} \\ \frac{1}{l^{\lambda_3}} \frac{\sigma_m^2}{\sigma_n^2} & \text{for the coefficients on cross-variable lags} \end{array} \right\}$$

and $\psi_{i,j}$ are assumed to follow a Gamma prior:

$$\psi_{i,j} \sim \mathcal{G}(\nu_\psi, \nu_\psi/2),$$

with an additional hyper-prior on $\nu_\psi \sim \mathcal{G}(1, 1)$. For $\sigma^2_{e_{ij}}$, priors are assumed to be $\sigma^2_{e_{ij}} \sim \mathcal{IG}(5, 0.04)$.

In Section 6.4, we also consider a version of model (B.1)-(B.2) with hierarchical shrinkage and Horseshoe prior. Similarly to the definitions of Section 3, let β_c, β_{AR}, and β_o be the coefficients related to intercept, own lags, and cross-variable lags, and let $\beta_{i,j}$ be the jth elements in the coefficient block i, where $i = \{c, AR, o\}$. In this case, we replace the prior specification in (B.4) by assuming that $\beta_{i,j}$ follows (10), where the global shrinkage parameter λ differs in each coefficient block.

Appendix B.2. Factor-augmented country-specific VARs

The factor-augmented country-specific VAR (CFAVAR) takes the form:

$$\begin{aligned} \begin{bmatrix} y_{i,t} \\ F_t \end{bmatrix} &= c_i + \sum_{l=1}^{p} B_{i,l} \begin{bmatrix} y_{i,t-l} \\ F_{t-l} \end{bmatrix} + u_{i,t} \\ Y_t^* &= \Lambda F_t + \varepsilon_t \\ F_t &= \sum_{l=1}^{q} \Pi_l F_{t-l} + v_t, v_t \sim i.i.d.\ N(0, \Sigma_v), \end{aligned}$$

[12]See, e.g., D'Agostino, Gambetti, and Giannone (2013) and Clark and Ravazzolo (2015).

where $Y_t^* = (y'_{1,t}, \ldots, y'_{i-1,t}, y'_{i+1,t}, y'_{N,t})'$ is the collection of foreign variables. F_t is an $r \times 1$ vector of weakly exogenous unobservable factors representing foreign information, which affect the variables in country i via the loadings $B^*_{i,l}, i = 1, \ldots, N, l = 1, \ldots, p$. Factors are estimated (recursively, as forecasting moves forward in time and the estimation sample expands) by principal components (see, e.g., Stock and Watson (2002a) and Stock and Watson (2002b)) and assumed to follow a VAR process with lag length q. In the VAR for $[y_{i,t}, F_t]$, the innovation vector $u_{i,t}$ includes the stochastic volatility structure previously indicated in the country-specific VAR's equation (B.2).

Priors for c_i and $B_{i,l}$ are specified in the same way as in the country-specific VARs. The same hyper-priors are imposed on (λ_1, λ_2), which are the overall tightness parameters on coefficients related to own lags and cross-variable lags. We specify the maximum number of factors and lag length to be $r^{\max} = 4$ and $q^{\max} = 4$, respectively. The number of factors is determined by the $IC2$ information criterion of Bai and Ng (2002), and the number of lags is determined by the Bayesian Information Criterion (BIC). The VAR for the factors is separately estimated by Bayesian methods with non-informative priors. Specifically, letting $\pi = vec([\Pi_1, \ldots, \Pi_q]')$, we specify $\pi \sim N(0, 100 \times I_{r^2q})$. Following Korobilis (2016), $\hat{\Sigma}_v$ is fixed at the OLS estimate to streamline computations (it also eliminates the uncertainty associated with covariance matrix estimation).

Appendix B.3. Global VARs

A GVAR model consists of a number of country-specific equations that are combined to form a global model. Assuming that the global economy consists of $N + 1$ countries, in the first step, we estimate the following country-specific VARX model for every country $i = 0, 1, ..., N$:

$$y_{i,t} = c_i + \sum_{l=1}^{p} B_{i,l} y_{i,t-l} + \sum_{l=0}^{p^*} B^*_{i,l} y^*_{i,t-l} + u_{i,t}, \tag{B.5}$$

$$u_{i,t} = A_i^{-1} H_{i,t}^{0.5} \epsilon_{i,t}, \quad \epsilon_{i,t} \overset{i.i.d.}{\sim} N(0, I_G), \tag{B.6}$$

where $t = 1, ..., T$, $y_{i,t}$ is a $G \times 1$ vector of endogenous variables in country i, c_i is a $G \times 1$ vector of intercept terms, $B_{i,l}(l = 1, ..., p)$ denotes the $G \times G$ matrix of parameters associated with lagged endogenous variables and $B^*_{i,l}(l = 0, 1, ..., p^*)$ is the matrix of parameters associated with contemporaneous and lagged weakly exogenous variables. The weakly exogenous foreign variables $y^*_{i,t}$ are constructed as a weighted average of the endogenous variables in other countries:

$$y^*_{i,t} = \sum_{j=0}^{N} w_{i,j} y_{j,t} \tag{B.7}$$

and the weights satisfy the following two restrictions: $w_{i,i} = 0$ and $\sum_{j=0}^{N} w_{i,j} = 1$. Weights are constructed from standardized bilateral trade flows. The data are available from Mohaddes and Raissi (2020).

In the second step, $N+1$ country-specific VARX models are stacked to form a global model, which is given by

$$Gy_t = c + \sum_{q=1}^{Q} H_q y_{t-q} + u_t, \tag{B.8}$$

where $y_t = (y'_{1,t}, ..., y'_{N,t})'$, $Q = \max(p, p^*)$, and G and H_q are both $NG \times NG$ dimensional coefficient matrices. Details on how to solve the global model can be found in Pesaran, Schuermann, and Smith (2009) and Huber (2016).

Priors for c_i and $B_{i,l}$ are specified in the same way as in the CFAVAR. More specifically, c_i and $B_{i,l}$ follow the same specification as in (B.4). For the prior on the elements of $B^*_{i,l}$, means are set to zero and variances are defined as: $\lambda_4 \frac{\sigma_m^2}{\sigma_n^2}$, where σ_m^2, σ_n^2 are obtained from univariate AR(1) regressions. We assume a Gamma prior for $\lambda_4 \sim \mathcal{G}(1, 0.02^2)$. Both p and p^* are set to 4.

Appendix B.4. Multi-country VARs

Appendix B.4.1. Factor shrinkage approach

The factor shrinkage approach used with the CC specification relies on the VAR written in system form. We define $X_t = I_{NG} \otimes x'_t$, where $x_t = (1, Y'_{t-1}, \ldots, Y'_{t-p})'$, β_i is the $k \times 1$ vector containing coefficients related to each i, $k = NGp + 1$, and $\beta = (\beta'_1, \ldots, \beta'_N)'$ is the $NGk \times 1$ vector containing all coefficients. Write the VAR as

$$Y_t = X_t \beta + u_t, \tag{B.9}$$

where $u_t \sim i.i.d.\ N(0, \Sigma_t)$.

Canova and Ciccarelli (2009) assume that the vector of coefficients β can be expressed as:

$$\beta = \sum_{i=1}^{F} \Xi_i \theta_i \tag{B.10}$$

where $\Xi = [\Xi_1, \ldots, \Xi_F]$ are known matrices and $\theta = (\theta'_1, \ldots, \theta'_F)'$ is a low dimensional vector ($\dim(\theta) < K$, where $K = kNG$) of unknown parameters, and $\theta_1, \ldots, \theta_F$ are mutually orthogonal.[13]

We consider the factorization used in Canova, Ciccarelli, and Ortega (2007) and Canova and Ciccarelli (2013). We assume $F = 4$. θ_1 is a scalar factor that is common across all countries, $\theta'_2 = (\theta_{2,1}, \ldots, \theta_{2,N})'$ is an $N \times 1$ vector of country-specific factors, $\theta'_3 = (\theta_{3,1}, \ldots, \theta_{3,G})'$ is a $G \times 1$ vector of variable-specific factors and $\theta'_4 = (\theta_{4,1}, \ldots, \theta_{4,p-1})'$ is a $(p-1) \times 1$ vector of lag-specific factors.[14] $\Xi_1, \ldots, \Xi_4$ are assumed to

[13]A more general form is $\beta = \sum_{i=1}^{F} \Xi_i \theta_i + e$, where $e \sim N(0, \Sigma \otimes \sigma^2 I)$ is an approximation error uncorrelated with u_t. However, most of the literature assumes an exact factorization ($\sigma^2 = 0$); see, for example, Canova, Ciccarelli, and Ortega (2007); Canova and Ciccarelli (2009); Dées and Güntner (2017). Koop and Korobilis (2019) estimate σ^2 by a forgetting factor approach and find that it is very small (< 0.01). In some limited checks, we found that considering the approximation error harms forecasting performance.

[14]To avoid collinearity with θ_1, θ_4 can contain at most $p-1$ components.

be known with elements associated with the corresponding original parameters equal to 1 and 0 otherwise. For example, consider a multi-country VAR model in (1) with $N = 2, G = 2, p = 1$. In this case, Ξ_1 is a 20×1 vector of ones, and Ξ_2 and Ξ_3 take the form:

$$\underset{20\times 2}{\Xi_2} = \begin{bmatrix} \iota_1 & 0 \\ \iota_1 & 0 \\ 0 & \iota_2 \\ 0 & \iota_2 \end{bmatrix}, \quad \underset{20\times 2}{\Xi_3} = \begin{bmatrix} \iota_3 & 0 \\ 0 & \iota_4 \\ \iota_3 & 0 \\ 0 & \iota_4 \end{bmatrix},$$

where $\iota_1 = (0, 1, 1, 0, 0)'$, $\iota_2 = (0, 0, 0, 1, 1)'$, $\iota_3 = (0, 1, 0, 1, 0)'$, and $\iota_4 = (0, 0, 1, 0, 1)'$. Thus, we can rewrite (B.9) as:

$$\begin{aligned} Y_t &= X_t\beta + u_t \\ &= X_t(\Xi\theta) + u_t = \tilde{X}_t\theta + u_t. \end{aligned} \tag{B.11}$$

In this case, $\dim(\theta) = N + G + p$. By construction, the $\tilde{X}_t$'s are linear combinations of the original right-hand-side variables in (B.9), and the parameterization above can capture comovement across lagged variables.

To incorporate SV, we decompose Σ_t as $\Sigma_t = A^{-1}H_tA'^{-1}$, where A is lower diagonal with diagonal elements equal to 1, and the diagonal elements in H_t evolve according to (B.3).

We specify the priors for θ, A, and Φ as (independent), Normal, Normal, and Inverse Wishart, respectively:

$$\theta \sim N(0, \underline{\Omega}_\theta), \quad a \sim N(0, \underline{\Omega}_a), \quad \Phi \sim IW(Q_0, W_0), \tag{B.12}$$

where a denotes the vector of free elements in A. The prior mean for θ is set to zero, and the prior covariance matrix $\underline{\Omega}_\theta$ is assumed to be diagonal. Letting $\omega_{\theta_{i,j}}$ be the elements in $\underline{\Omega}_\theta$ associated with the jth elements in θ_i, where $i = 1, \ldots, 4$, then

$$\underline{\omega}_{\theta_{i,j}} = \left\{ \begin{array}{ll} \sum_{m=1}^{NG} \sigma_m^2 & i = 1, 2, 3 \\ \frac{\sum_{m=1}^{NG} \sigma_m^2}{l^2}, & i = 4, l = 2, \ldots, p \end{array} \right\}.$$

The prior mean for a is set to 0, and the prior variance is set to $\underline{\Omega}_a = 10 \times I$. Q_0, W_0 are specified as $Q_0 = NG + 2$, $W_0 = 0.01 \times I$.

Appendix B.4.2. Prior specifications for other models

For the approach in Angelini, et al. (2019) and the hierarchical shrinkage considered in this paper, the prior for free elements in A is assumed to be Normal with zero mean and variance equal to $10 \times I_{NG}$. The prior for Φ takes the form $\Phi \sim IW(Q_0, W_0)$, and Q_0, W_0 are specified as $Q_0 = NG + 2$, $W_0 = 0.01 \times I$.

For the prior in (B.4), σ_i^2, σ_j^2 are obtained from univariate AR(1) regressions. The prior for the intercept is assumed to be uninformative by setting the prior variance equal to $100 \times \sigma_i^2$, where σ_i^2 is again from a univariate AR(1) regression. The hyper-priors on overall shrinkage parameters are specified in the same way as in country-specific VARs. For the additional hyperparameter λ_4 controlling the tightness for coefficients related to cross-variable lags for foreign countries, we use a prior of $\lambda_4 \sim \mathcal{G}(1, 0.02^2)$.

For the SSSS prior, we follow Korobilis (2016) exactly. For (7) and (8), we set $\xi_{ij}^2 = \tau_{ij}^2 = 4$ and $\underline{c}^{\text{DI}} = \underline{c}^{\text{CSH}} = 0.0025$. The priors for indicators are specified as

$$
\begin{aligned}
\gamma_{ij}^{\text{DI}} &\sim Bernoulli(\pi_{ij}^{\text{DI}}), \;\; \pi_{ij}^{\text{DI}} \sim \mathcal{B}(1,1) \\
\gamma_{ij}^{\text{CSH}} &\sim Bernoulli(\pi_{ij}^{\text{CSH}}), \;\; \pi_{ij}^{\text{CSH}} \sim \mathcal{B}(1,1).
\end{aligned}
$$

For the Horseshoe prior, no more prior specifications are needed. For the Normal-Gamma prior, recall that we specify $a^{\omega} \sim \mathcal{E}(b)$ and $\kappa^2 \sim \mathcal{G}(d_1, d_2)$. We set b equal to the number of coefficients in each block and elicit a non-informative prior for κ^2 by setting $d_1 = d_2 = 0.01$. For the Normal-Gamma-Gamma prior, recall that $2a \sim \mathcal{B}(\alpha_a, \beta_a)$, $2c \sim \mathcal{B}(\alpha_c, \beta_c)$, and we set $\alpha_a = \alpha_c = 2, \beta_a = \beta_c = 1$.

Appendix B.5. Univariate models

For AR(p)-SV models applied to each scalar output growth or interest rate variable, generally denoted y_t, we have

$$
\begin{aligned}
y_t &= c + \sum_{\ell=1}^{p} \rho_\ell y_{t-\ell} + u_t, \\
u_t &= h_t^{0.5} v_t, \;\; v_t \overset{i.i.d.}{\sim} N(0,1), \\
\log h_t &= \log h_{t-1} + e_t, \;\; e_t \overset{i.i.d.}{\sim} N(0, \sigma_e^2).
\end{aligned}
$$

As in Clark and Ravazzolo (2015), lag length is set to 2 for output growth and 4 for the interest rate. Letting $\theta = (c, \rho_1, \ldots, \rho_p)'$, we specify the following priors:

$$
\theta \sim N(0, V), \;\; \sigma_e^2 \sim IG(v_h, S_h), \;\; \log h_0 \sim N(a_0, b_0).
$$

V is assumed to be diagonal with elements equal to $\frac{\theta_1}{\ell^{\theta_2}}$, $\ell = 1, \ldots, p$, for autoregressive coefficients and $100 \times \hat{\sigma}_y^2$ for the intercept. θ_1 is set to 0.04, θ_2 is set to 2, and $\hat{\sigma}_y^2$ is obtained from a univariate AR(1) regression. We use a modestly informative prior for σ_e^2 to control the time variation by setting v_h equal to 2 and S_h to 0.04. For the prior on initial conditions, we set $a_0 = 0$ and $b_0 = 10$.

For the UCSV model, we have

$$y_t = \tau_t + \varepsilon_t^y, \quad \varepsilon_t^y \sim N(0, e^{h_t}),$$
$$\tau_t = \tau_{t-1} + \varepsilon_t^\tau, \quad \varepsilon_t^\tau \sim N(0, e^{g_t}),$$
$$h_t = h_{t-1} + \varepsilon_t^h, \quad \varepsilon_t^h \sim N(0, \omega_h^2),$$
$$g_t = g_{t-1} + \varepsilon_t^g, \quad \varepsilon_t^g \sim N(0, \omega_g^2),$$

with initial conditions τ_0, h_0 and g_0 as unknown parameters. We can rewrite the above UCSV model in the non-centered parameterization:

$$y_t = \tau_t + e^{\frac{1}{2}(h_0+\omega_h \tilde{h}_t)}\tilde{\varepsilon}_t^y,$$
$$\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0+\omega_g \tilde{g}_t)}\tilde{\varepsilon}_t^\tau,$$
$$\tilde{h}_t = \tilde{h}_{t-1} + \tilde{\varepsilon}_t^h,$$
$$\tilde{g}_t = \tilde{g}_{t-1} + \tilde{\varepsilon}_t^g,$$

where $\tilde{h}_0 = \tilde{g}_0 = 0$ and $\tilde{\varepsilon}_t^y$, $\tilde{\varepsilon}_t^\tau$, $\tilde{\varepsilon}_t^h$, and $\tilde{\varepsilon}_t^g$ are all *i.i.d.* $N(0, 1)$. We assume Normal priors for all model parameters: $\omega_h \sim N(0, 0.2^2)$, $\omega_g \sim N(0, 0.2^2)$, $h_0 \sim N(0, 10)$, $g_0 \sim N(0, 10)$, and $\tau_0 \sim N(0, 10)$.

Appendix C. Algorithms

Appendix C.1. Algorithms for VARs with Minnesota-type prior

For all the country-specific VARs, country-specific factor-augmented VARs, global VARs, and multi-country VARs with Minnesota prior, the MCMC samplers follow almost exactly the steps in Carriero, et al. (2022), but an additional step is needed to update prior tightness parameters. We highlight three issues related to the sampler, and refer interested readers to Appendix A.3 in their paper for other details.

Step 1: Update $\beta|\cdot$. We update the coefficients equation by equation, as in the corrected triangular algorithm in Carriero, et al. (2022). Details can be found in Appendix C.5.

Step 2: Update $\lambda_i|\cdot$, $i = 1, 2, 4$. Let S_{λ_i}, $i = 1, 2, 4$, be the collection of all indexes such that parameters associated with the overall shrinkage parameters belong to this set. It can easily be shown that, with a Gamma prior, $\lambda_i \sim \mathcal{G}(1, c_i)$, conditional posteriors follow a Generalized Inverse Gaussian distribution:

$$\lambda_i|\cdot \sim \mathcal{GIG}\Big(1 - \frac{\dim(S_{\lambda_i})}{2}, 2c_i, \sum_{(i,j)\in S_{\lambda_i}} \frac{\beta_{i,j}^2}{2C_{i,j}}\Big).$$

The density of $x \sim \mathcal{GIG}(p, a, b)$ is given by $f(x) \propto x^{p-1} \exp\left(-(ax+b/x)/2\right)$. dim denotes the dimension of the set, and $C_{i,j}$ are the prior local variance parameters (the elements in (B.4) without an overall shrinkage parameter).

Step 3: Update the volatility. For the volatility estimation, let $\tilde{u}_t = Au_t$ denote the rescaled residuals. The elements of $\tilde{u}_t$ obey the following process:

$$\ln \tilde{u}^2_{ij,t} = \ln h_{ij,t} + \ln \epsilon^2_{ij,t}, \quad i = 1, \ldots, N, j = 1, \ldots, G.$$

So, together with state equation (B.2), we have a non-linear and non-Gaussian state space system. To get the volatility estimates, we use the KSC algorithm, first introduced in Kim, Shephard, and Chib (1998) and detailed for VAR models in Del Negro and Primiceri (2015). We use a 10-state mixture of Normals to approximate the distribution of non-Gaussian errors $\ln \epsilon^2_{ij,t}$. The details of approximation are provided in Table 1 of Omori, et al. (2007).

Step 4: Update the free elements in A. This can be done with the equation-by-equation approach of Cogley and Sargent (2005) or with the joint approach of Chan (2017). For the latter, letting a denote the free elements in A, it can be shown that a can be interpreted as the coefficients from the regression:

$$u_t = K_t a + e_t, e_t \sim N(0, D_t),$$

where $D_t = diag(h_{1,t}, \ldots, h_{NG,t})$, and K_t is given as

$$K_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ -u_{1t} & 0 & 0 & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & -u_{1t} & -u_{2t} & 0 & 0 & \cdots & \cdots & \vdots \\ \vdots & & & \ddots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & -u_{1t} & \cdots & -u_{(NG-1)t} \end{bmatrix}.$$

This permits drawing a jointly. Given the prior $a \sim N(0, \underline{\Omega}_a)$, the posterior is also Gaussian $a|\beta, h, \Phi, Y \sim N(\overline{\mu}_a, \overline{\Omega}_a)$, where

$$\overline{\Omega}_a = (\underline{\Omega}_a^{-1} + K'H^{-1}K)^{-1}$$
$$\overline{\mu}_a = \overline{\Omega}_a K'H^{-1}u.$$

This algorithm can be more efficient than the equation-by-equation approach, because a is updated jointly. However, the band matrix K_t does not have a fixed bandwidth (the number of non-zeros elements increases with model size). Thus, letting n denote the number of variables in the model, the complexity of this algorithm is still $O(n^3)$, and the estimation quickly becomes computationally demanding as the model size increases. Accordingly, for country-specific models, which are small ($n = N = 3$), we use this algorithm to update a. But for multi-country models, which are large ($n = NG = 21$), we use the algorithm of Cogley and Sargent (2005) to draw a equation by equation.

Step 5: Update $\Phi|\cdot$. Since we elicit a conditionally conjugate prior, the conditional posterior takes the same form, which can be shown to be:

$$\Phi|\cdot \sim IW\Big(Q_0 + T, W_0 + \sum_{t=1}^{T}\Big(\log(h_t) - \log(h_{t-1})\Big)\Big(\log(h_t) - \log(h_{t-1})\Big)'\Big).$$

Appendix C.2. Algorithm for multi-country VAR with factor shrinkage

Most of the steps of the algorithm for the CC specification follow from the previous section, except that we have to adapt step 1's treatment of the VAR's coefficients. With the transformation, we see that given $\theta \sim N(0, \underline{\Omega}_\theta)$, the conditional posterior $\theta|Y, a, h, \Phi$ is multivariate Normal, $N(\overline{\mu}_\theta, \overline{\Omega}_\theta)$, with moments:

$$\overline{\Omega}_\theta = (\underline{\Omega}_\theta^{-1} + \tilde{Z}'\tilde{\Sigma}^{-1}\tilde{Z})^{-1}$$
$$\overline{\mu}_\theta = \overline{\Omega}_\theta \tilde{Z}'\tilde{\Sigma}^{-1} Y,$$

where $Y, \tilde{Z}$ are stacked versions of $Y_t, \tilde{Z}_t$ and $\tilde{\Sigma} = diag(\Sigma_1, \ldots, \Sigma_T)$.

Appendix C.3. Algorithms for multi-country VARs with hierarchical shrinkage

As in Algorithm 1 in the main text, the MCMC estimation involves 5 steps. The only new step compared to above is to update the prior variance parameters and associated hyperparameters. We provide details of the conditional posterior distributions for these parameters. In Section 6.4, we also estimate country-specific VAR-SV specifications with hierarchical shrinkage and Horseshoe prior. The algorithm follows exactly the ones described below.

First, consider the Horseshoe prior:

$$\beta_j|\omega_j^2 \sim N(0, \omega_j^2), \;\; \omega_j^2|\gamma_j^2 \sim \mathcal{G}(\frac{1}{2}, \gamma_j^2), \;\; \gamma_j^2 \sim \mathcal{G}(\frac{1}{2}, \lambda),$$

and $\lambda \sim C^+(0, 1)$. It follows from straightforward calculation that

$$\omega_j^2|\cdot \sim \mathcal{GIG}(0, 2\gamma_j^2, \beta_j^2),$$

where $\mathcal{GIG}(p, a, b)$ denotes the Generalized Inverse Gaussian distribution with pdf given by $f(x) \propto x^{p-1}\exp\big(-(ax + b/x)/2\big)$. For the conditional posterior of $\gamma_j^2|\cdot$, since the Gamma distribution is conjugate for the Gamma likelihood function, we have that

$$\gamma_j^2|\cdot \sim \mathcal{G}(1, \lambda + \omega_j^2).$$

Updating $\lambda|\cdot$ follows the same steps as above since the prior admits the hierarchical representation: $\lambda \sim \mathcal{G}(\frac{1}{2}, \xi^2)$, $\xi^2 \sim \mathcal{G}(\frac{1}{2}, 1)$.

Second, consider the Normal-Gamma prior:

$$\beta_j|\omega_j^2 \sim N(0, \omega_j^2), \;\; \omega_j^2 \sim \mathcal{G}(a^\omega, \frac{a^\omega \kappa^2}{2}),$$

and $a^\omega \sim \mathcal{E}(b)$ and $\kappa^2 \sim \mathcal{G}(d_1, d_2)$. It follows similarly as in the Horseshoe prior that

$$\omega_j^2|\cdot \sim \mathcal{GIG}(a^\omega - 0.5, a^\omega \kappa^2, \beta_j^2).$$

The conditional posterior for $a^\omega|\cdot$ is not available in closed form. We use adaptive Random Walk Metropolis-Hastings algorithms as in Roberts and Rosenthal (2009) with acceptance probability given by

$$\min\left\{1, \frac{p(a^{\omega,\text{new}})a^{\omega,\text{new}}}{p(a^\omega)a^\omega} \prod_j \frac{p(\beta_j|a^{\omega,\text{new}}, \kappa^2)}{p(\beta_j|a^\omega, \kappa^2)}\right\},$$

where the marginal prior is given by

$$p(\beta_j|a^\omega, \kappa^2) = \frac{(\sqrt{a^\omega \kappa^2})^{a^\omega + \frac{1}{2}}}{\sqrt{\pi} 2^{a^\omega - \frac{1}{2}} \Gamma(a^\omega)} \left|\beta_j\right|^{a^\omega - \frac{1}{2}} K_{a^\omega - \frac{1}{2}}\left(\sqrt{a^\omega \kappa^2}\left|\beta_j\right|\right),$$

and $K(\cdot)$ denotes a modified Bessel function of the second kind. At each iteration i, a new value $a^{\omega,\text{new}}$ is proposed according to

$$\log a^{\omega,\text{new}} = \log a^\omega + \epsilon_j, \quad \epsilon_j \sim N(0, \sigma_{\psi_j}^{2(i)}). \tag{C.1}$$

The variance of the increments is fixed at 1 for the first 50 iterations, and then updated by

$$\log \sigma_{a^\omega}^{2(i+1)} = \log \sigma_{a^\omega}^{2(i)} + \frac{1}{i^q}(\hat{\alpha} - \alpha^*), \tag{C.2}$$

where $\hat{\alpha}$ is the estimated acceptance probability of current draws and α^* is the desired acceptance probability. The parameter q controls the degree of vanishing adaption, which is necessary to make the adaptive algorithm valid.[15] This algorithm leads to an average acceptance rate that converges to α^*. Following Griffin and Brown (2017), we set $q = 0.55, \alpha^* = 0.3$. Then updating $\kappa^2|\cdot$ is quite straightforward since it again follows a Gamma distribution:

$$\kappa^2|\cdot \sim \mathcal{G}\left(Ma^\omega + d_1, d_2 + a^\omega \sum_j \omega_j^2\right),$$

[15]This means that the variances of increments are fixed as $i \to \infty$. Two conditions are provided in equations (1.1) and (1.2) of Roberts and Rosenthal (2009). The condition in equation (1.2) in their paper is generally satisfied provided that ψ is bounded above.

where M denotes the number of parameters in each block.

Finally, consider the Normal-Gamma-Gamma prior:

$$\beta_j|\tau_j^2, \lambda_j^2 \sim N\Big(0, \phi\frac{\tau_j^2}{\lambda_j^2}\Big), \;\; \tau_j^2 \sim \mathcal{G}(a, 1), \;\; \lambda_j^2 \sim \mathcal{G}(c, 1),$$

where $\phi = 2c/(a\kappa^2)$, $2a \sim \mathcal{B}(\alpha_a, \beta_a)$, $2c \sim \mathcal{B}(\alpha_c, \beta_c)$, and $\kappa^2|a, c \sim F(2a, 2c)$. We proceed as in Cadonna, Frühwirth-Schnatter, and Knaus (2020). As we use marginalized distributions in each step to improve sampling efficiency, the steps described below are not interchangeable.

Step a: Update $a|\cdot$. Use the prior $p(\beta_j|\lambda_j^2, a, c)$, marginalized *w.r.t.* τ_j^2, to draw $a|\cdot$ via an adaptive Random Walk Metropolis-Hastings algorithm on $z = \log(a/(0.5 - a))$. The variance of the increments is updated as in the Normal-Gamma case. At each iteration m, letting a^* be the candidate draw and $a^{(m-1)}$ be the previous draw, the acceptance probability is given by

$$\min\left\{1, \frac{q_a(a^*)}{q_a(a^{(m-1)})}\right\}, \quad q_a(a) = p(a|\cdot)a(0.5 - a).$$

Letting m be the number of parameters in each block, $\log q_a(a)$ is given by

$$\begin{aligned}
\log q_a(a) = a\Big(&-m\log 2 + \frac{m}{2}\log\kappa^2 - \frac{m}{2}\log c + \frac{1}{2}\sum_j \log\lambda_j^2 + \frac{1}{2}\sum_j \log\beta_j^2\Big)\\
&+ \frac{5}{4}m\log a + m\frac{a}{2}\log a - m\log\Gamma(a+1)\\
&+ \sum_j \log K_{a-\frac{1}{2}}\Big(\beta_j\sqrt{\lambda_j^2\kappa^2 a/c}\Big)\\
&- \log\mathcal{B}(a, c) + a\Big(\log a + \log(\frac{\kappa^2}{2c})\Big) - \log a - (a + c)\log\Big(1 + \frac{a\kappa^2}{2c}\Big)\\
&+ (\alpha_a - 1)\log(2a) - (\beta_a - 1)\log(1 - 2a)\\
&+ \log a + \log(0.5 - a).
\end{aligned}$$

Step b: Update $\tau_j^2|\cdot$. This step is simple, as the conditional posterior is again $\mathcal{GIG}$:

$$\tau_j^2|\cdot \sim \mathcal{GIG}\Big(a - \frac{1}{2}, 2, \frac{\lambda_j^2\beta_j^2}{\phi}\Big).$$

Step c: Update $c|\cdot$. Use the prior $p(\beta_j|\tau_j^2, a, c)$, marginalized *w.r.t.* λ_j^2, to draw $c|\cdot$ via an adaptive Random Walk Metropolis-Hastings algorithm on $z = \log(c/(0.5 - c))$. The variance of the increments is updated as in the Normal-Gamma case. At each iteration m, letting c^* be the candidate draw and $c^{(m-1)}$ be the previous draw, the acceptance probability is given by

$$\min\left\{1, \frac{q_c(c^*)}{q_c(c^{(m-1)})}\right\}, \quad q_c(c) = p(c|\cdot)c(0.5 - c).$$

Letting m be the number of parameters in each block, $\log q_c(c)$ is given by

$$\begin{aligned}\log q_c(c) &= m \log \Gamma(c + 0.5) - m \log \Gamma(c + 1) + \frac{m}{2} \log c \\ &- (c + 0.5)\Big(\sum_j \log(4c\tau_j^2 + \beta_j^2\kappa^2 a) - \sum_j \log(4c\tau_j^2)\Big) \\ &- \log \mathcal{B}(a, c) - (a - 1)\log c - (a + c)\log(1 + \frac{a\kappa^2}{2c}) \\ &+ (\alpha_c - 1)\log(2c) + (\beta_c - 1)\log(1 - 2c) \\ &+ \log c + \log(0.5 - c).\end{aligned}$$

Step d: Update $\lambda_j^2|\cdot$. This step is simple; the conditional posterior is $\mathcal{G}$:

$$\lambda_j^2|\cdot \sim \mathcal{G}\Big(\frac{1}{2} + c, \frac{\beta_j^2}{2\phi\tau_j^2} + 1\Big).$$

Step e: Update $\kappa^2|\cdot$. Notice that the prior of κ^2 admits the following hierarchical representation: $\kappa^2|a \sim \mathcal{G}(a, d_2)$, $d_2|a, c \sim \mathcal{G}(c, \frac{2c}{a})$. Then updating $\kappa^2|\cdot$ involves first sampling from

$$d_2|\cdot \sim \mathcal{G}(a + c, \kappa^2 + \frac{2c}{a}),$$

then sampling from (m is the number of parameters in each block)

$$\kappa^2|\cdot \sim \mathcal{G}(\frac{m}{2} + a, \frac{a}{4c}\sum_j \frac{\lambda_j^2}{\tau_j^2}\beta_j^2 + d_2).$$

Appendix C.4. Corrected triangular algorithm

Consider an n-variable reduced-form $VAR(p)$ model as in Carriero, et al. (2022):

$$y_t = \Pi' x_t + A^{-1}\Lambda_t^{0.5}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, I_n),$$

where $t = 1, \ldots, T$, x_t is an $(np + 1) \times 1$ dimensional vector containing the lags of y_t and an intercept, $\Pi = (\Pi_0, \Pi_1, \ldots, \Pi_p)'$ is an $(np + 1) \times n$ matrix of coefficients, A^{-1} is a unit lower triangular matrix, and $\Lambda_t^{0.5}$ is diagonal with the log of the generic j-th element following a random walk process.

Defining $\tilde{y}_t = Ay_t$ with generic j-th element $\tilde{y}_{j,t} = y_{j,t} + a_{j,1}y_{1,t} + \cdots + a_{j,j-1}y_{j-1,t}$, consider the triangular representation of the system:

$$\tilde{y}_t = A\Pi' x_t + \Lambda_t^{0.5}\epsilon_t = A(x_t'\Pi)' + \Lambda_t^{0.5}\epsilon_t,$$

which can be expressed as the following system of equations:

$$\begin{aligned}
\tilde{y}_{1,t} &= x_t'\pi^{(1)} + \lambda_{1,t}^{0.5}\epsilon_{1,t} \\
\tilde{y}_{2,t} &= a_{2,1}x_t'\pi^{(1)} + x_t'\pi^{(2)} + \lambda_{2,t}^{0.5}\epsilon_{2,t} \\
\tilde{y}_{3,t} &= a_{3,1}x_t'\pi^{(1)} + a_{3,2}x_t'\pi^{(2)} + x_t'\pi^{(3)} + \lambda_{3,t}^{0.5}\epsilon_{3,t} \\
&\vdots \\
\tilde{y}_{n,t} &= a_{n,1}x_t'\pi^{(1)} + \cdots + a_{n,n-1}x_t'\pi^{(n-1)} + x_t'\pi^{(n)} + \lambda_{n,t}^{0.5}\epsilon_{n,t},
\end{aligned}$$

where $\pi^{(j)}$ denotes the coefficients of the j-th equation. Clearly, $\pi^{(j)}$ appears not only in equation j but also in equations $j+1$ through n. Letting $z_{j+l,t} = \tilde{y}_{j+l,t} - \sum_{i\neq j,i=1}^{j+l} a_{j+l,i}x_t'\pi^{(i)}$, for $l = 0, ..., n-j$, and $a_{i,i} = 1$, consider the following system of equations:

$$\begin{aligned}
z_{j,t} &= x_t'\pi^{(j)} + \lambda_{j,t}^{0.5}\epsilon_{j,t} \\
z_{j+1,t} &= a_{j+1,j}x_t'\pi^{(j)} + \lambda_{j+1,t}^{0.5}\epsilon_{j+1,t} \\
&\vdots \\
z_{n,t} &= a_{n,j}x_t'\pi^{(j)} + \lambda_{n,t}^{0.5}\epsilon_{n,t}.
\end{aligned}$$

Then, using the above triangular representation, the full conditional posterior of $\pi^{(j)}|\cdot$ follows immediately from standard Bayesian linear regression results (assuming that prior means are zero):

$$\pi^{(j)}|\cdot \sim N\Big(\overline{\mu}_{\pi^{(j)}}, \overline{\Omega}_{\pi^{(j)}}\Big),$$

where

$$\begin{aligned}
\overline{\Omega}_{\pi^{(j)}}^{-1} &= \underline{\Omega}_{\pi^{(j)}}^{-1} + \sum_{i=j}^{n} a_{i,j}^2 \sum_{t=1}^{T} \frac{1}{\lambda_{i,t}} x_t x_t' \\
\overline{\mu}_{\pi^{(j)}} &= \overline{\Omega}_{\pi^{(j)}} \times \Big(\sum_{i=j}^{n} a_{i,j} \sum_{t=1}^{T} \frac{1}{\lambda_{i,t}} x_t z_{i,t} \Big),
\end{aligned}$$

with $a_{i,i} = 1$.

Appendix C.5. Algorithms for SSSS prior

The algorithms described in Appendix A.3 of Korobilis (2016) can be easily extended to our case with SV. Only step 1 has to be modified. In particular, let $Y = (y_1 \ \cdots \ y_T)'$, $x_t = (1, y'_{t-1})'$, and $X = (x_1 \ \cdots \ x_T)'$, and write the model as

$$Y = XB + U,$$

where $U = (u_1 \ \cdots \ u_T)'$. The sampler involves the following steps:

Step a: Update $\text{vec}(B)|\cdot$. It can be shown that

$$\text{vec}(B)|\cdot \sim N(\Gamma \times \mu_B, D_B),$$

where

$$D_B = \Big(V + \sum_{t=1}^{T} (\Sigma_t^{-1} \otimes x_t' x_t)\Big)^{-1}, \ \ \mu_B = D_B\Big(\text{vec}(\sum_{t=1}^{T} x_t y_t' \Sigma_t^{-1})\Big).$$

The diagonal matrix V contains prior variances; details of constructing the indicator matrix Γ can be found in Korobilis (2016).

Steps b,c,d,e: These follow exactly as in steps 2,3,4,5 in Korobilis (2016).

Steps f,g,h: Update free elements in A, stochastic volatility, and related parameters. These steps follow the corresponding steps used for the multi-country VAR with the Minnesota-type prior.

Appendix C.6. Algorithms for country-specific VAR with hierarchical shrinkage

We follow exactly the algorithms described in Chan (2021). Estimation for the intercept, autoregressive coefficients, free elements in A, and stochastic volatility is very similar to the algorithms used in this paper. It is worth mentioning that, as in Chan (2021), the model has been first transformed to structural form, and then estimation is performed equation by equation. For hyperparameters related to the Normal-Gamma prior, since a slightly different parameterization is used there, the updating of hyperparameters is slightly different. The conditional posterior for $\psi_{i,j}|\cdot$ is also $\mathcal{GIG}$, but with a slightly different parameterization. An independent Metropolis-Hastings algorithm is used to update $\nu_\psi|\cdot$. We refer the reader to Section 4 and Appendix B in that paper for more details.

Appendix C.7. Algorithms for univariate models

We use the algorithms as described in Clark and Ravazzolo (2015) to estimate AR(p)-SV models. The steps to draw intercept and autoregressive parameters follow from standard linear regression results. To estimate stochastic volatility and related parameters, we follow the procedures described in Section 7.1 in Chan (2017). For the UCSV model, we estimate it in non-centered parameterization and then transform back to the centered parameterization to perform predictive simulation. Estimation details can be found in Appendix B in Chan (2018) and in Section 7.2 in Chan (2017).

Appendix D. Additional empirical results

Table D.1: Comparison of HS-CSH and baseline HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.229	-1.999	-0.623	-1.236	-1.836	-2.761	Mean	-0.204	0.087	-1.018	-0.759	0.610	0.933
Median	-1.053	-2.029	-0.610	-1.183	-2.209	-3.580	Median	0.325	0.264	0.287	0.077	0.349	0.639
Min	-6.251	-7.402	-3.926	-5.022	-6.251	-7.402	Min	-8.772	-7.086	-8.772	-7.086	-6.593	-5.164
Max	3.282	3.078	2.128	2.284	3.282	3.078	Max	10.556	10.025	3.519	4.223	10.556	10.025
% > 0	32.143	28.571	35.714	28.571	28.571	28.571	% > 0	57.143	55.952	54.762	52.381	59.524	59.524
$\%p <= 0.05$	0	2.381	0	2.381	0	2.381	$\%p <= 0.05$	0	1.190	0	2.381	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	7.951	6.192	3.960	2.191	11.942	10.193							
Median	6.169	5.192	3.180	1.804	10.788	8.174							
Min	-4.416	-5.618	-4.416	-5.618	0.337	0.383							
Max	28.668	23.514	21.308	14.492	28.668	23.514							
% > 0	90.476	86.905	80.952	73.810	100	100							
$\%p <= 0.05$	8.333	10.714	2.381	2.381	14.286	19.048							

Notes: "HS-CSH" is the multi-country VAR model in which all the parameters related to CSH restrictions follow the same Horseshoe prior specification. The table provides summary statistics for the performance of this alternative model compared to the multi-country HS specification. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.2: Comparison of HS-A and baseline HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.592	-3.194	-1.386	-2.642	-1.798	-3.746	Mean	-2.180	-1.956	-2.282	-2.190	-2.079	-1.721
Median	-1.626	-3.109	-1.160	-2.488	-2.434	-4.376	Median	-2.037	-1.909	-1.173	-1.290	-2.533	-3.450
Min	-7.806	-12.719	-7.806	-10.600	-7.287	-12.719	Min	-12.187	-10.419	-12.187	-10.419	-9.568	-8.276
Max	4.116	3.688	2.804	3.197	4.116	3.688	Max	9.258	10.280	3.404	4.276	9.258	10.280
% > 0	33.333	28.571	33.333	28.571	33.333	28.571	% > 0	28.571	25	28.571	21.429	28.571	28.571
$\%p <= 0.05$	1.190	4.762	2.381	4.762	0	4.762	$\%p <= 0.05$	0	0	0	0	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	10.697	10.033	4.370	3.699	17.023	16.367							
Median	9.769	10.128	4.496	3.532	17.011	16.445							
Min	-7.071	-10.052	-7.071	-10.052	6.479	7.464							
Max	25.761	26.416	17.567	15.201	25.761	26.416							
% > 0	90.476	86.905	80.952	73.810	100	100							
$\%p <= 0.05$	21.429	38.095	9.524	23.810	33.333	52.381							

Notes: "HS-A" is the multi-country VAR model in which all the parameters follow the same Horseshoe prior specification. The table provides summary statistics for the performance of this alternative model compared to the multi-country HS specification. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.3: Comparison of HS-E and baseline HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.154	-2.286	-1.212	-2.101	-1.095	-2.470	Mean	-1.756	-2.309	-2.332	-2.680	-1.181	-1.939
Median	-1.454	-2.461	-1.105	-1.768	-1.608	-2.767	Median	-1.109	-2.087	-1.109	-2.370	-1.036	-1.991
Min	-7.244	-10.132	-7.244	-9.228	-6.265	-10.132	Min	-13.790	-15.398	-13.790	-15.398	-9.156	-10.805
Max	4.161	3.987	3.126	3.642	4.161	3.987	Max	11.063	9.459	3.316	4.634	11.063	9.459
% > 0	33.333	28.571	38.095	28.571	28.571	28.571	% > 0	26.190	26.190	28.571	30.952	23.810	21.429
%p <= 0.05	0	1.190	0	2.381	0	0	%p <= 0.05	0	3.571	0	7.143	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	11.280	10.700	4.059	3.313	18.501	18.087							
Median	10.568	10.984	2.596	2.770	16.977	17.372							
Min	-7.634	-11.672	-7.634	-11.672	6.737	7.870							
Max	34.537	28.239	25.968	19.593	34.537	28.239							
% > 0	83.333	82.143	66.667	64.286	100	100							
%p <= 0.05	13.095	20.238	2.381	9.524	23.810	30.952							

Notes: "HS-E" is the multi-country VAR model in which all the parameters in each equation follow the same Horseshoe prior specification. The table provides summary statistics for the performance of this alternative model compared to the multi-country HS specification. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.4: Comparison of Horseshoe priors with and without SV: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.852	-3.445	-0.893	-2.525	-2.811	-4.364	Mean	-22.974	-19.720	-11.820	-12.712	-34.127	-26.727
Median	-1.962	-3.434	-1.123	-2.603	-2.722	-4.309	Median	-16.394	-15.737	-7.538	-8.186	-24.002	-19.492
Min	-9.097	-11.549	-6.032	-8.837	-9.097	-11.549	Min	-98.999	-69.770	-56.606	-52.922	-98.999	-69.770
Max	5.197	4.580	5.197	4.580	2.362	0.835	Max	4.018	0.548	4.018	0.548	-7.184	-7.248
%> 0	28.571	19.048	38.095	28.571	19.048	9.524	%> 0	4.762	1.190	9.524	2.381	0	0
%p <= 0.05	3.571	5.952	0	2.381	7.143	9.524	%p <= 0.05	15.476	15.476	16.667	16.667	14.286	14.286
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	-9.429	-13.515	-12.149	-20.332	-6.709	-6.698							
Median	-7.592	-6.062	-8.375	-14.525	-7.590	-2.534							
Min	-57.608	-73.394	-57.608	-73.394	-27.100	-47.395							
Max	10.402	14.686	9.970	11.987	10.402	14.686							
%> 0	26.190	28.571	21.429	14.286	30.952	42.857							
%p <= 0.05	11.905	28.571	16.667	42.857	7.143	14.286							

Notes: The table provides summary statistics for the performance of the alternative model with SV compared to the multi-country HS specification with SV. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.5: Comparison of Horseshoe priors with expanding versus rolling windows: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	1.095	0.342	1.229	0.707	0.960	-0.023	Mean	3.268	1.872	3.989	2.046	2.546	1.698
Median	0.849	0.160	1.102	0.528	0.759	0.002	Median	3.006	1.788	3.533	2.128	2.742	1.208
Min	-1.632	-1.930	-1.632	-1.930	-0.688	-1.772	Min	-1.479	-3.063	-1.318	-2.326	-1.479	-3.063
Max	7.256	4.464	7.256	4.464	3.948	3.023	Max	10.905	5.850	10.905	5.495	7.575	5.850
%> 0	83.333	54.762	80.952	59.524	85.714	50	%> 0	92.857	85.714	95.238	88.095	90.476	83.333
$\%p$ <= 0.05	3.571	4.762	7.143	9.524	0	0	$\%p$ <= 0.05	0	0	0	0	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	1.928	1.330	0.539	-0.077	3.317	2.738							
Median	1.435	0.461	-0.338	0.043	3.237	1.713							
Min	-6.350	-6.095	-6.350	-6.095	-2.769	-4.450							
Max	12.168	12.503	9.226	9.878	12.168	12.503							
%> 0	59.524	57.143	47.619	50	71.429	64.286							
$\%p$ <= 0.05	10.714	4.762	7.143	0	14.286	9.524							

Notes: The table provides summary statistics for the performance of the HS model estimated with a rolling approach relative to the paper's baseline recursive approach. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.6: Comparison with univariate models with HS baseline featuring SV: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	0.535	0.732	0.290	0.396	0.779	1.067	Mean	1.279	3.874	0.252	0.927	2.306	6.822
Median	-0.513	-0.461	-0.342	-0.406	-0.750	-0.744	Median	2.022	3.073	1.740	1.213	2.572	7.117
Min	-3.642	-4.655	-3.616	-3.813	-3.642	-4.655	Min	-13.484	-7.475	-13.484	-7.475	-11.329	-3.670
Max	7.571	10.119	5.646	6.972	7.571	10.119	Max	13.247	17.693	6.875	7.681	13.247	17.693
%> 0	38.095	40.476	40.476	42.857	35.714	38.095	%> 0	70.238	77.381	59.524	64.286	80.952	90.476
$\%p$ <= 0.05	13.095	14.286	11.905	7.143	14.286	21.429	$\%p$ <= 0.05	0	8.333	0	0	0	16.667
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	8.982	7.799	8.442	7.370	9.522	8.227							
Median	10.504	10.207	8.651	8.779	14.056	16.663							
Min	-19.264	-28.167	-15.896	-22.757	-19.264	-28.167							
Max	28.698	29.027	24.232	25.927	28.698	29.027							
%> 0	78.571	63.095	85.714	69.048	71.429	57.143							
$\%p$ <= 0.05	11.905	25	9.524	23.810	14.286	26.190							

Notes: This table presents descriptive statistics on comparisons of forecasting performance for the multi-country VAR-SV model with the Horseshoe prior (the paper's HS specification) relative to univariate models with SV. For output growth and the interest rate, we use an AR(p)-SV model, with $p = 2$ for output growth and $p = 4$ for the interest rate. For inflation, we use an unobserved component model with SV, as in Chan (2018). Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

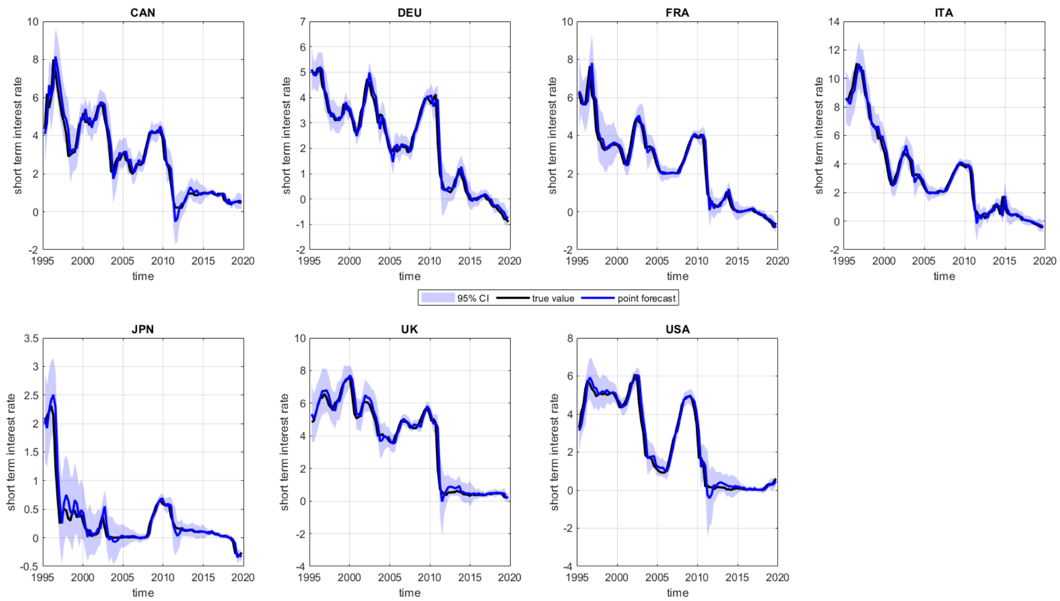

Figure D.1: The figures present 1-step-ahead short-term interest rate forecasts for all G7 countries. The blue line and shaded areas are point forecasts and associated 95 percent forecast intervals. The black line shows the true values.

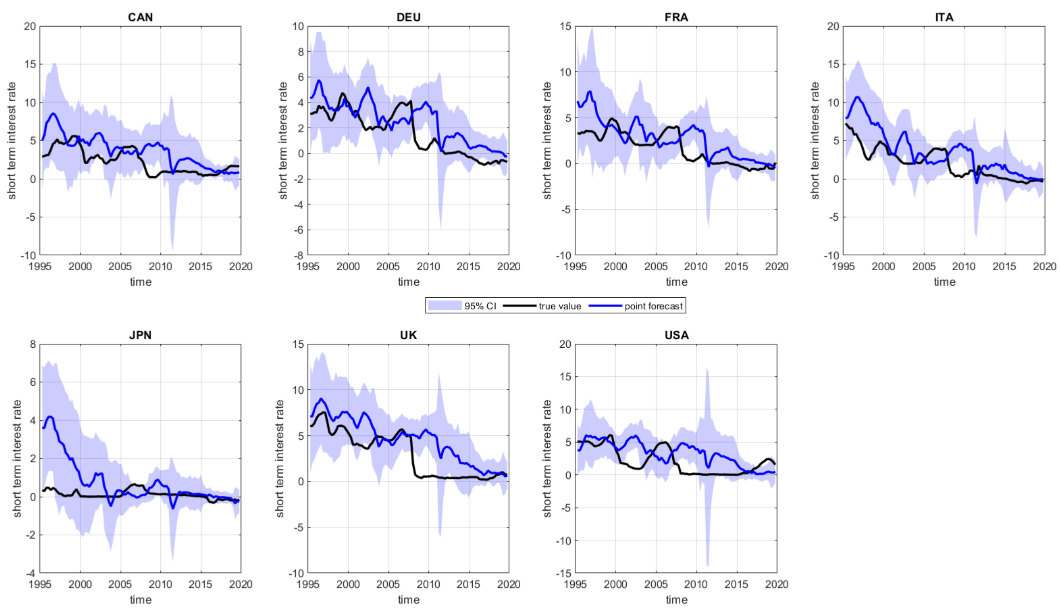

Figure D.2: The figures present 12-steps-ahead short-term interest rate forecasts for all G7 countries. The blue line and shaded areas are point forecasts and associated 95 percent forecast intervals. The black line shows the true values.

Table D.7: Directional forecast: 1-step-ahead changes in output growth

	CAN	DEU	FRA	ITA	JPN	UK	USA
HS	-0.380	6.715	2.508	5.870	7.556	8.174	9.225
	(0.648)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)
CVAR	-2.183	0.421	-2.928	-2.307	-2.318	-1.668	-3.797
	(0.986)	(0.337)	(0.998)	(0.990)	(0.990)	(0.952)	(0.999)

Notes: This table presents test statistics and associated p-values for directional predictive performance of 1-step-ahead changes in output growth from multi-country VAR-SV model with Horseshoe prior and single-country VAR-SV benchmark. The test statistics are computed according to equation (6) in Pesaran and Timmermann (1992).

Table D.8: Comparison with hierarchical country-specific VARs featuring SV: HS prior versus Chan (2021), descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	1.778	2.238	1.525	1.924	2.032	2.552	Mean	2.944	3.462	1.520	2.027	4.369	4.896
Median	1.947	1.876	1.412	1.737	1.951	2.556	Median	2.174	2.881	0.924	2.124	4.364	4.747
Min	-3.041	-3.547	-3.041	-3.547	-1.975	-0.621	Min	-2.281	-1.249	-2.281	-1.249	-1.583	-0.724
Max	6.223	6.509	6.223	6.509	5.940	6.200	Max	9.802	10.208	6.099	6.701	9.802	10.208
%> 0	79.762	84.524	83.333	85.714	76.190	83.333	%> 0	80.952	86.905	76.190	83.333	85.714	90.476
$\%p <= 0.05$	17.857	14.286	11.905	11.905	23.810	16.667	$\%p <= 0.05$	1.190	22.619	2.381	11.905	0	33.333
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	9.936	9.228	8.174	7.606	11.699	10.851							
Median	11.195	10.659	8.529	8.511	14.053	13.219							
Min	-0.939	-0.742	-0.939	-0.742	1.082	-0.365							
Max	20.896	21.690	20.896	17.718	20.696	21.690							
%> 0	97.619	90.476	95.238	95.238	100	85.714							
$\%p <= 0.05$	30.952	30.952	11.905	14.286	50	47.619							

Notes: This table presents descriptive statistics on comparisons of forecasting performance for the hierarchical shrinkage in the country-specific VAR-SV model with the Horseshoe prior (the paper's HS specification) relative to the hierarchical shrinkage with the Normal-Gamma prior as in Chan (2021). Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.9: Comparison of country-specific VAR-SV and baseline multi-country VAR-SV with HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	1.134	1.341	1.537	1.758	0.731	0.924	Mean	1.429	2.286	0.825	1.189	2.034	3.382
Median	0.775	1.152	1.064	1.201	0.617	0.968	Median	1.964	2.600	1.951	2.344	1.964	3.497
Min	-4.553	-3.130	-4.553	-2.116	-2.354	-3.130	Min	-13.569	-10.703	-13.569	-10.703	-11.219	-9.106
Max	5.461	5.745	5.461	5.745	3.937	5.032	Max	13.926	15.588	7.518	8.836	13.926	15.588
%> 0	72.619	73.810	85.714	80.952	59.524	66.667	%> 0	78.571	75	73.810	69.048	83.333	80.952
%p <= 0.05	3.571	8.333	7.143	16.667	0	0	%p <= 0.05	4.762	15.476	0	4.762	9.524	26.190
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	1.627	1.764	-0.253	0.435	3.506	3.093							
Median	1.030	3.074	-1.630	2.123	3.445	4.184							
Min	-12.774	-16.290	-11.779	-15.435	-12.774	-16.290							
Max	19.525	20.786	18.605	20.046	19.525	20.786							
%> 0	55.952	60.714	38.095	57.143	73.810	64.286							
%p <= 0.05	23.810	33.333	7.143	23.810	40.476	42.857							

Notes: This table presents descriptive statistics on comparisons of forecasting performance for the country-specific VAR-SV model and multi-country VAR-SV model with the Horseshoe prior (the paper's HS specification). Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone (2020).

Table D.10: Loss function levels for the benchmark CVAR specification

	RMSFE				CRPS			
Output growth	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 1$	$h = 4$	$h = 8$	$h = 12$
CAN	2.319	2.712	2.637	2.647	1.245	1.466	1.419	1.402
DEU	3.592	3.595	3.569	3.470	1.832	1.805	1.798	1.764
FRA	1.630	2.068	2.116	2.144	0.892	1.115	1.134	1.149
ITA	2.539	2.985	2.964	2.935	1.335	1.580	1.553	1.515
JPN	4.185	4.167	4.156	4.251	2.208	2.187	2.159	2.261
UK	2.070	2.542	2.509	2.534	1.080	1.311	1.282	1.287
USA	2.335	2.548	2.594	2.542	1.266	1.364	1.393	1.367
Inflation	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 1$	$h = 4$	$h = 8$	$h = 12$
CAN	1.870	1.722	1.816	1.809	0.998	1.006	1.060	1.085
DEU	1.140	1.277	1.377	1.376	0.663	0.743	0.814	0.794
FRA	1.118	1.410	1.439	1.456	0.618	0.783	0.848	0.875
ITA	0.934	1.498	1.690	1.777	0.503	0.830	0.946	1.005
JPN	1.652	1.797	1.846	1.858	0.891	0.978	1.015	1.023
UK	0.982	1.215	1.384	1.358	0.542	0.701	0.778	0.813
USA	2.151	2.227	2.223	2.193	0.988	1.102	1.185	1.160
Interest rate	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 1$	$h = 4$	$h = 8$	$h = 12$
CAN	0.474	1.327	2.130	2.588	0.231	0.704	1.187	1.502
DEU	0.323	1.068	1.795	2.202	0.162	0.589	1.083	1.374
FRA	0.416	1.298	2.083	2.445	0.192	0.675	1.182	1.419
ITA	0.461	1.502	2.518	3.190	0.234	0.777	1.393	1.817
JPN	0.177	0.740	1.342	1.577	0.068	0.278	0.536	0.676
UK	0.418	1.233	1.884	2.289	0.189	0.630	1.012	1.295
USA	0.353	1.188	2.076	2.644	0.173	0.648	1.205	1.588