Forecasting with Global Vector Autoregressive Models: A Bayesian Approach*

Jesús Crespo Cuaresma[a,b,c,d], Martin Feldkircher[e], and Florian Huber[a,e]

[a]Vienna University of Economics and Business (WU)
[b]Wittgenstein Centre for Demography and Human Capital (WIC)
[c]International Institute for Applied Systems Analysis (IIASA)
[d]Austrian Institute of Economic Research (WIFO)
[e]Oesterreichische Nationalbank (OeNB)

Abstract

This paper develops a Bayesian variant of global vector autoregressive (B-GVAR) models to forecast an international set of macroeconomic and financial variables. We propose a set of hierarchical priors and compare the predictive performance of B-GVAR models in terms of point and density forecasts for one-quarter-ahead and four-quarters-ahead forecast horizons. We find that forecasts can be improved by employing a global framework and hierarchical priors which induce country-specific degrees of shrinkage on the coefficients of the GVAR model. Forecasts from various B-GVAR specifications tend to outperform forecasts from a naive univariate model, a global model without shrinkage on the parameters and country-specific vector autoregressions.

JEL Classification: *C11, C32, C53, C55, F44.*
Keywords: *International forecasts, shrinkage priors, GVAR.*

*The opinions expressed in this paper are those of the authors and do not necessarily reflect the official viewpoint of the OeNB or of the Eurosystem. We would like to thank two anonymous referees, Marta Bańbura, Jonas Dovern, Domenico Giannone, Sylvia Frühwirth-Schnatter, Blazej Mazur, Luca Onorante, M. Hashem Pesaran, Philipp Piribauer and the participants of the 8th ECB Workshop on Forecasting Techniques, Frankfurt, the first annual conference of the International Association for Applied Econometrics, Queen Mary University, London, the annual meeting of the Austrian Economic Association (NOeG) and internal research seminars at the Vienna University of Economics and Business and the OeNB for helpful comments. Email: jcrespo@wu.ac.at, martin.feldkircher@oenb.at and florian.huber@oenb.at.

1 Introduction

The rise in international trade and cross-border financial flows in recent decades implies that countries are more than ever exposed to economic shocks from abroad, as demonstrated by the recent global financial crisis. Hence, macroeconomic tools that treat countries as isolated from the rest of the world may miss important information for forecasting and counterfactual analysis. Such concerns do not arise with global vector autoregressive (GVAR) models, as they accommodate spillovers from the global economy in a systematic and transparent manner. The GVAR framework consists of single-country models that are stacked to yield a comprehensive representation of the world economy.

The empirical literature on GVAR models has been largely influenced by the work of M. Hashem Pesaran and co-authors (Pesaran *et al.*, 2004; Garrat *et al.*, 2006; Dees *et al.*, 2007a;b). Recent papers have advanced the literature on GVAR modeling in terms of country coverage (Feldkircher, 2015), identification of shocks (Eickmeier & Ng, 2015) and the specification of international linkages (Chudik & Fratzscher, 2011; Eickmeier & Ng, 2015; Feldkircher & Huber, 2015; Galesi & Sgherri, 2013). Most of the existing applications of GVAR models concentrate on the quantitative assessment of the propagation of macroeconomic shocks using historical data, while very few contributions have addressed their forecasting performance. Pesaran *et al.* (2009) provide an out-of-sample forecasting exercise and conclude that taking global links across economies into account using GVAR models leads to more accurate predictions than using forecasts based on univariate specifications for output and inflation. Employing a GVAR model to forecast macroeconomic variables in five Asian economies, Han & Ng (2011) find that one-step-ahead forecasts from GVAR models outperform those of stand-alone VAR specifications for short-term interest rates and real equity prices, while Greenwood-Nimmo *et al.* (2012) confirm the superiority of GVAR specifications over univariate benchmark models at forecast horizons beyond four quarters.

As an alternative to the GVAR framework a related strand of the literature advocates the estimation of large VARs or panel VARs using Bayesian techniques. More specifically, Bańbura *et al.* (2010) assess the forecasting performance of a large-scale monetary VAR based on more than 100 macroeconomic variables and sectoral information. They show that forecasts of these large-scale models can outperform small benchmark VARs when the degree of shrinkage on the parameters is set in relation to the size of the model. Giannone & Reichlin (2009) and Alessi & Bańbura (2009) propose to exploit these shrinkage properties and estimate Bayesian VARs with a large cross-section of countries. Alessi & Bańbura (2009) show that Bayesian VAR specifications as well as dynamic factor models are able to yield accurate one-quarter to four-quarters-ahead forecasts for international macroeconomic data. Koop & Korobilis (2015) propose a panel VAR framework that overcomes the problem of overparametrization by averaging over different restrictions on interdependencies between and heterogeneities across cross-sectional units. More recently, Korobilis (2015) advocate a particular class of priors that allows for soft clustering of variables or countries, arguing that classical shrinkage priors are inappropriate for panel VARs.

In this contribution, we propose using established shrinkage priors and develop a Bayesian GVAR (B-GVAR) model. Akin to the GVAR framework, we assume that links among economies are determined exogenously, while we borrow strength from the Bayesian literature in estimating the individual country models. This allows us to keep the virtues of the GVAR framework with regard to offering a coherent way for policy and counterfactual analysis. Our model includes standard variables that are often employed in small-country VARs such as output, inflation, short-term and long-term interest rates, the real exchange rate, equity prices and the oil price as a global control variable (see e.g., Dees *et al.*, 2007b;a; Pesaran *et al.*, 2004; 2009, among others). This set of variables is extended to feature total credit (domestic and cross-border credit), which can act as an important transmission channel of international shocks.

We compare forecasts of the B-GVAR model under prior specifications that resurface frequently in Bayesian VAR empirical studies: the conjugate Minnesota prior (Litterman, 1986) and its version with a fixed (non-stochastic) variance-covariance error structure, and a weighted average of a Minnesota type prior, the "initial dummy observation" prior, which accommodates potential cointegration relationships among the variables considered, and the "sum-of-coefficients" prior, which facilitates soft-differencing (Doan *et al.*, 1984; Sims, 1992; Sims & Zha, 1998). We extend this set of random-walk priors to include the stochastic search variable selection (SSVS) prior proposed by George *et al.* (2008) for VAR models. Since the hyperparameters for all priors are elicited locally (i.e., for the country model), our approach induces country-specific degrees of shrinkage on the parameters, which is expected to improve forecasts significantly. B-GVAR models are thus expected to be less prone to overfitting (Giannone & Reichlin, 2009) and allow the researcher to include prior beliefs in the model, while still taking the long-run co-movement of variables into account. We compare our battery of priors using an expanding window to forecast developments one-quarter-ahead and four-quarters-ahead. These forecasts are benchmarked to forecasts of a fifth-order autoregressive model with drift term by means of root mean squared errors for point forecasts, and log predictive scores for density forecasts. As another competitor, and to assess the importance of international linkages for forecasting, we evaluate forecasts from isolated, country-specific Bayesian vector autoregressions.

Our analysis provides several new insights on the specification and estimation of global macroeconomic models. First, we find that forecasts can be improved by employing a global framework that allows for country-specific degrees of shrinkage on the parameters. The proposed Bayesian specifications of the GVAR tend to improve upon forecasts from the naive model, a global model without shrinkage and a shrinkage model that neglects international linkages. Second, we find that the prior specification put forward in Sims & Zha (1998), the fixed-covariance Minnesota prior and the SSVS prior all show a strong forecasting performance. The latter outperforms other priors systematically in terms of density forecasts. Third, our analysis indicates that Latin American variables are particularly hard to forecast, while the forecast performance for developed economies is more homogeneous among the specifications considered.

The paper is structured as follows. Section 2 provides a brief description of the global VAR model, while Section 3 derives its Bayesian variant. In Section 4 we present the data and perform the forecast evaluation exercise. Finally, Section 5 concludes.

2 The GVAR Model

GVAR specifications constitute a compact representation of the world economy designed to model multilateral dependencies among economies across the globe. Basically, a GVAR model consists of a number of country-specific specifications that are combined to form a global model.

The first step is to estimate separate multivariate time series models. In our case, these are standard vector autoregressive models involving a set of endogenous variables and enlarged by weakly exogenous and global control variables (VARX* model). Assuming that our global economy consists of $N+1$ countries, we estimate a VARX* of the following form for every country $i = 0, ..., N$,

$$x_{it} = a_{i0} + \sum_{s=1}^{p} \Phi_{is} x_{it-s} + \sum_{r=0}^{p^*} \Lambda_{ir} x^*_{it-r} + \varepsilon_{it}, \tag{2.1}$$

where x_{it} is a $k_i \times 1$ vector of endogenous variables in country i at time $t \in 1, ..., T$, a_{i0} is a k_i-dimensional vector of intercept terms, Φ_{is} $(s = 1, \dots, p)$ denotes the $k_i \times k_i$ matrix of parameters associated with the lagged endogenous variables and Λ_{ir} $(r = 1, \dots, p^*)$ are the coefficient matrices of the k_i^* weakly exogenous variables, which are of dimension $k_i \times k_i^*$. Furthermore, ε_{it} is the standard zero-mean vector error term with variance-covariance matrix $\Sigma_{\varepsilon i}$.

The weakly exogenous or *foreign* variables, x^*_{it}, are constructed as a weighted average of the endogenous variables in other economies,

$$x^*_{it} = \sum_{j=0}^{N} \omega_{ij} x_{jt}, \tag{2.2}$$

with ω_{ij} denoting the (non-negative) weight corresponding to the pair of country i and country j. We assume that $\omega_{ii} = 0$ and $\sum_{j=0}^{N} \omega_{ij} = 1$. The weights ω_{ij} reflect economic and financial ties among economies, which are usually approximated using data on (standardized) bilateral trade flows.[1] The assumption that the x^*_{it} variables are weakly exogenous at the individual level reflects the belief that most countries are small relative to the world economy.

Following Pesaran *et al.* (2004) we stack the $N+1$ country-specific models to obtain a global model, which is given by

$$Gx_t = a_0 + \sum_{q=1}^{Q} H_q x_{t-q} + \epsilon_t. \tag{2.3}$$

Here, G is a $k \times k$-dimensional matrix that establishes contemporaneous relations between countries, with $k = \sum_{i=0}^{N} k_i$. Furthermore, let a_0 be a k-dimensional vector associated with the constant and $H_q (q = 1, \dots, Q)$ is a $k \times k$-dimensional global coefficient matrix (with $Q = \max(p, p^*)$). The matrices G, a_0 and H_q are complex functions of the corresponding country-specific parameters and the bilateral weights. Finally, ϵ_t is a global vector error term with variance-covariance matrix Σ_ϵ. Further details on the derivation of the GVAR model can be found in Appendix B.

3 The B-GVAR: Priors over Parameters

Bayesian analysis of the GVAR model requires the elicitation of prior distributions for all parameters of the model. We use several prior structures that have been developed for VAR specifications over the parameters of the individual country-specific models, which we extend to account for the presence of (weakly) exogenous variables.[2] For prior implementation, it proves convenient to rewrite the model in (2.1) as

$$x_{it} = \Pi_i' Z_{it-1} + \varepsilon_{it}, \tag{3.1}$$

where $Z_{it-1} = (1, x'_{it-1}, \dots, x'_{it-p}, x^{*'}_{it}, \dots, x^{*'}_{it-p^*})'$ is of dimension $K_i \times 1$, where $K_i = 1 + k_i p + k_i^*(p^* + 1)$ and $\Pi_i = (a_{i0}, \Phi_{i1}, \dots, \Phi_{ip}, \Lambda_{i0}, \dots, \Lambda_{ip^*})'$ denotes a $K_i \times k_i$ matrix of stacked coefficients. Up to this point we have not adopted any distributional assumptions for ε_{it}. We complete the model specification by assuming that the errors ε_{it} are multivariate Gaussian, i.e., $\varepsilon_{it} \sim \mathcal{N}(0, \Sigma_{\varepsilon i})$.

Rewriting the model in terms of full-data matrices yields

$$x_i = Z_i \Pi_i + \varepsilon_i \tag{3.2}$$

[1] See e.g., Eickmeier & Ng (2015) and Feldkircher & Huber (2015) for an application using a broad set of different weights.

[2] Karlsson (2012) provides an excellent overview for Bayesian VAR models.

where x_i is a $T \times k_i$ matrix of stacked endogenous variables, Z_i is a $T \times K_i$ matrix of stacked explanatory variables and ε_i is a $T \times k_i$ matrix of errors. Furthermore, let $\Psi_i = \text{vec}(\Pi_i)$ denote the v_i-dimensional coefficient vector with $v_i = k_i K_i$.

The General Conjugate Prior Setup

We start with the simplest prior for the coefficients of the country-specific VARX* models, which is the natural conjugate prior. In the VARX* framework, we impose an inverted Wishart prior on $\Sigma_{\varepsilon i}$ and a multivariate Gaussian prior on Ψ_i

$$\Psi_i | \Sigma_{\varepsilon i} \sim \mathcal{N}(\underline{\Psi}_i, \Sigma_{\varepsilon i} \otimes \underline{V}_i), \tag{3.3}$$

$$\Sigma_{\varepsilon i} \sim \mathcal{IW}(\underline{S}_i, \underline{v}_i), \tag{3.4}$$

where $\underline{\Psi}_i$ and $\underline{V}_i$ denote prior mean and variance, respectively. Additionally, we let $\underline{S}_i$ denote the prior scale matrix and $\underline{v}_i$ the prior degrees of freedom for $\Sigma_{\varepsilon i}$. The use of such a natural conjugate prior allows us to exploit a Kronecker factorization of the likelihood due to the prior dependence between Ψ_i and $\Sigma_{\varepsilon i}$, which translates into significant computational advantages. Especially for forecasting applications where the model has to be re-estimated several times over a training sample, this proves to be a significant advantage. However, it is worth noting that the Kronecker factorization implies prior variances on the coefficients that are proportional across equations of the country model, which might be very restrictive.

Following the literature on Bayesian VARs (Litterman, 1986; Sims, 1992; Sims & Zha, 1998), the most common choices for $\underline{\Psi}_i$ and $\underline{V}_i$ are given by the so-called random walk priors. Under the prior, the variables in the system are assumed to follow simple random walks. To implement this prior, we set the prior mean according to

$$\underline{\Psi}_{ij} = \begin{cases} \underline{a}_{ij} \text{ for the first own lag of endogenous variable } j \text{ in equation } j \\ 0 \text{ in all other cases.} \end{cases} \tag{3.5}$$

where $\underline{a}_{ij}$ $(j = 1, \ldots, k_i)$ refers to the prior mean over the parameter associated with the first own lag of the k_i endogenous variables. These are set to one for variables in levels, leading to the traditional random walk prior. The assumption that the endogenous variables a priori follow random walk processes at the local level directly carries over to the global model. To see this, note that under the prior model, the coefficients associated with the contemporaneous and lagged (weakly) exogenous variables are set equal to zero. Moreover, the coefficients corresponding to higher lag orders of the endogenous variables are also set equal to zero. Hence, the G and H matrices reduce to $k \times k$ identity matrices. Consequently, the global prior model is given by

$$x_t = x_{t-1} + e_t, \tag{3.6}$$

where it is straightforward to show that the variance-covariance matrix of e_t is a block-diagonal matrix with the corresponding ith block being equal to the prior expectation of $\Sigma_{\varepsilon i}$. The only assumption which is crucial for this result to hold is that the prior mean of coefficients related to the weakly exogenous variables is set to zero.

Several choices are recommended in the literature for the elicitation of $\underline{V}_i$, which translate into different assumptions about the behavior of the prior model. Doan *et al.* (1984), Kadiyala & Karlsson (1997) and Sims & Zha (1998) propose three prominent prior specifications that have been frequently employed by practitioners. The most prominent prior is the Minnesota prior, which has a proven track record in terms of forecasting performance. The Minnesota prior specifies the prior variance on the coefficients, $\underline{V}_i$, such that the parameter corresponding to lag r of variable g is given by

$$\underline{V}_{ig,r} = \begin{cases} \frac{\alpha_{i1}}{r^{\kappa}\sigma_{ig}} & \text{for the coefficient of the } r\text{th lag of variable } g \\ \frac{\alpha_{i2}}{(1+r)^{\kappa}\sigma^{*}_{ig}} & \text{for the coefficient of the } r\text{th lag of variable } g \text{ if weakly exogenous} \\ \alpha_{i3} & \text{for the deterministic part of the model.} \end{cases} \quad (3.7)$$

Here, hyperparameters α_{i1} and α_{i2} control the tightness of the prior on the endogenous and weakly exogenous part, respectively. Moreover, the priors are scaled using standard deviations obtained by running univariate autoregressions on the particular variables. Specifically, σ_{ig} refers to the standard deviation of a univariate autoregressive model for the corresponding variable, whereas σ^{*}_{ig} denotes the standard deviation obtained from an autoregressive model of the gth weakly exogenous variable. Finally, r^{κ} is a deterministic function of the lag length.[3] Consequently, the strength of the prior belief in the random walk specification is governed by α_1. The hyperparameter κ increasingly tightens the variance on the prior for distant lags, reflecting the belief that longer lags of the variables are more likely have zero coefficients.

There is a direct link between the locally specified $\underline{V}_i$ and the global specification of the GVAR model. It is straightforward to show that there exists a relationship between the prior variances on the weakly exogenous variables and the variances related to other countries' endogenous variables (termed global prior variances). As an illustration, consider β^{n}_{i1}, the coefficient associated with the first lag of the nth weakly exogenous variable, i.e., $x^{n*}_{it} = \sum_{j=0}^{N}\omega_{ij}x^{n}_{jt}$, with prior variance given by $\underline{\sigma}^2_{in}$. Then, $\beta^{n}_{i1,j} = \omega_{ij}\beta^{n}_{i1}$ denotes country i's coefficient corresponding to the nth variable of country j with (prior) variance given by $\omega^2_{ij}\underline{\sigma}^2_{in}$. Hence, the corresponding global prior variance is simply scaled down by the trade links between countries i and j.[4] Note that in contrast to a clean Bayesian approach this

[3] In order to achieve a symmetric specification of endogenous and weakly exogenous variables in the variance prior, the term is $(1+r)^{\kappa}$ in the expression corresponding to the latter, since contemporaneous weakly exogenous variables are included in the model.

[4] A more formal treatment can be found in Appendix D

implies that the corresponding weights are not treated as being random and thus we do not integrate out uncertainty surrounding the cross-country linkages. Moreover, in a fully Bayesian framework the prior variances at the global level would be proportional to each other. An alternative approach that provides an agnostic and statistical measure of connectivity between countries and where fully conjugate priors are readily available is the Bayesian panel VAR, see Koop & Korobilis (2015) or Korobilis (2015).

The Minnesota prior can be implemented by means of so-called dummy observations. Following Bańbura *et al.* (2010) and Koop (2013), the moments of the conjugate Minnesota prior can be matched attaching the following set of artificial dummy observations to the actual data

$$\underline{x}_i^M = \begin{pmatrix} 0_{1\times k_i} \\ \cdots\cdots \\ \mathrm{diag}(\underline{a}_{i1}\sigma_{i1},\ldots,\underline{a}_{ik_i}\sigma_{ik_i})/\alpha_{i1} \\ 0_{k_i(p-1)\times k_i} \\ \cdots\cdots \\ 0_{k_i^*(p^*+1)\times k_i} \\ \cdots\cdots \\ \mathrm{diag}(\sigma_{i1},\ldots,\sigma_{ik_i}) \end{pmatrix}, \underline{Z}_i^M = \begin{pmatrix} \frac{1}{\alpha_3} & & 0_{1\times(K_i-1)} \\ \cdots & \cdots & \cdots \\ 0_{k_ip\times 1} & J_p\otimes \mathrm{diag}(\sigma_{i1},\ldots,\sigma_{ik_i})/\alpha_1 & 0_{k_ip\times(K_i-k_ip-1)} \\ \cdots & \cdots & \cdots \\ 0_{(k_i^*(p^*+1))\times(k_ip+1)} & & J_{p^*}\otimes\mathrm{diag}(\sigma_{i1}^*,\ldots,\sigma_{ik_i^*}^*)/\alpha_{i2} \\ \cdots & \cdots & \cdots \\ & 0_{k_i\times K_i} & \end{pmatrix}, \tag{3.8}$$

where $J_p = \mathrm{diag}(1,2,...,p)$ and $J_{p*} = \mathrm{diag}(1,2,...,p+1)$. Additionally, 0_{nq} denotes an $n\times q$ dimensional matrix consisting exclusively of zeros. This setting corresponds to the normal inverse Wishart prior proposed by Kadiyala & Karlsson (1997).

The first block of $\underline{x}_i$ and $\underline{Z}_i$ implements the prior on the deterministic part of the model whereas the second block implements the random walk prior. Finally, the last two blocks implement the priors on the weakly exogenous variables and $\Sigma_{\varepsilon i}$, respectively.

Second, we consider the "sum-of-coefficients" prior, which softly forces the posterior distribution towards a specification in first differences. This implies that coefficients associated with own, lagged variables in each equation should sum to unity while other coefficients are being pushed towards zero. Implementation of this prior is straightforward by adding the following set of dummy observations to the data

$$\begin{aligned}\underline{x}_i^S &= \left(\mathrm{diag}(\underline{a}_{i1}\underline{\mu}_{i1},\ldots,\underline{a}_{ik_i}\underline{\mu}_{ik_i})/\theta_{i1},\right)\\ \underline{Z}_i^S &= \left(0_{k_i\times 1} \quad \iota_{1\times p}\otimes\mathrm{diag}(\underline{a}_{i1}\underline{\mu}_{i1},\ldots,\underline{a}_{ik_i}\underline{\mu}_{ik_i})/\theta_{i1} \quad 0_{k_i\times(k_i^*(p^*+1))}\right),\end{aligned} \tag{3.9}$$

where $\underline{\mu}_{ij}$ $(j=1,\ldots,k_i)$ denotes the pre-sample mean of the endogenous variables usually calculated by using the first p observations, $\iota_{1\times p}$ is a p-dimensional row vector of ones and θ_{i1} is a country-specific hyperparameter controlling the tightness of the prior.

The fact that this prior is not consistent with cointegration gives rise to the "dummy-initial-observation" prior. This prior pushes variables in a country-specific VAR towards their unconditional (stationary) mean, or toward a situation where there is at least one unit root present. That is, either the process has a unit root, or it is stationary and starts near its mean, implying a penalty for models with inherent initial transient dynamics (Sims, 1992). Implementation boils down to attaching the following set of dummy observations to the actual data.

$$\begin{aligned} \underline{x}_i^I &= \left(\frac{(\underline{\mu}_{i1},\ldots,\underline{\mu}_{ik_i})}{\theta_{i2}}\right), \\ \underline{Z}_i^I &= \left(0 \quad \iota_{1\times p} \otimes \frac{(\underline{\mu}_{i1},\ldots,\underline{\mu}_{ik_i})}{\theta_{i2}} \quad \iota_{1\times p^*} \otimes \frac{(\underline{\mu}^*_{i1},\ldots,\underline{\mu}^*_{ik_i})}{\theta_{i2}}\right). \end{aligned} \tag{3.10}$$

$\underline{\mu}^*_{ij}$ $(j = 1,\ldots,k_i^*)$ denote pre-sample averages from the weakly exogenous variables and θ_{i2} is a hyperparameter controlling the tightness of the dummy-initial-observation prior.

In practice, macroeconomists usually incorporate all three versions of the random walk prior, which can be implemented in a straightforward fashion by combining the three pairs of dummy observations given in equations (3.8) to (3.10). The final prior, as motivated in Sims & Zha (1998), is then simply a weighted average of the three individual priors described above, where the weights attached to each prior are determined by the associated hyperparameters. Several studies have emphasized the usefulness of such a weighted prior structure (Bańbura *et al.*, 2010; Giannone *et al.*, 2013).

Natural conjugate priors require prior dependence between $\Sigma_{\varepsilon i}$ and Ψ_i. The traditional implementation of the Minnesota prior drops this dependence, which provides more flexibility in terms of prior elicitation. In the empirical application in subsection 4.3 we also consider a variant of the Minnesota prior where the variance-covariance matrix of the error term is considered non-random. This prior simply replaces the posterior of $\Sigma_{\varepsilon i}$ by a known estimate $\hat{\Sigma}_{\varepsilon i}$, which leads to analytical posterior solutions.

Stochastic Search Variable Selection (SSVS) Prior

The conjugate priors discussed above apply by definition the same degree of shrinkage across equations. It might be appealing to provide more flexibility in the specification of the prior variance-covariance matrix on the coefficients and move away from the random walk prior model.

The SSVS prior, put forward by George & McCulloch (1993) and subsequently introduced to the VAR literature by George *et al.* (2008), imposes a mixture of Normal distributions on each coefficient of the VARX* which is usually specified as

$$\Psi_{ij}|\delta_{ij} \sim (1-\delta_{ij})\mathcal{N}(0,\tau_{0j}^2) + \delta_{ij}\mathcal{N}(0,\tau_{1j}^2). \tag{3.11}$$

Here, δ_{ij} is a binary random variable corresponding to coefficient j in country model i. It equals one if the corresponding variable is included in the model and zero if it is a priori excluded from the respective country model. The Normal distribution corresponding to $\delta_{ij} = 0$ is typically specified with τ_{0j}^2 close to zero, which pushes the respective coefficient towards zero. The prior variance of the Normal distribution for $\delta_{ij} = 1$, τ_{1j}^2, is set to a comparatively large value implying a relatively uninformative prior on coefficient j conditional on inclusion. The prior mean of the normal distributions in equation (3.11), routinely chosen to be zero in applications of the SSVS prior, may of course be centered around some other value $\underline{\Psi}_{ij}$. The SSVS prior has been recently applied within a GVAR framework in Feldkircher & Huber (2015) to examine the international dimension of US economic shocks.

Defining a scalar parameter d_{ij} such that

$$d_{ij} = \begin{cases} \tau_{0j} \text{ if } \delta_{ij} = 1 \\ \tau_{1j} \text{ if } \delta_{ij} = 0 \end{cases} \tag{3.12}$$

and collecting all d_{ij} ($j = 1, ..., v_i$) in a $v_i \times v_i$ matrix $D_i = \text{diag}(d_{i1}, \ldots, d_{iv_i})$ the prior on Ψ_i simply reduces to the following hierarchical prior setup

$$\Psi_i | D_i \sim \mathcal{N}(0, \underline{R}_i), \tag{3.13}$$

$$\Sigma_{\varepsilon i} \sim \mathcal{IW}(\underline{S}_i, \underline{v}_i), \tag{3.14}$$

where $\underline{R}_i = D_i D_i$ and the prior on $\Sigma_{\varepsilon i}$ is a standard inverse Wishart prior with prior degrees of freedom given by $\underline{v}_i$ and prior scale matrix $\underline{S}_i$. Note that the lack of prior dependence between $\Sigma_{\varepsilon i}$ and Ψ_i renders this prior (even conditionally) non-conjugate.

Finally, we follow George *et al.* (2008) and impose a Bernoulli prior on δ_{ij}

$$\delta_{ij} \sim \text{Bernoulli}(\underline{q}_{ij}), \tag{3.15}$$

where $\underline{q}_{ij}$ denotes the prior inclusion probability of variable j in country i.

Estimation of this model requires Markov Chain Monte Carlo (MCMC) methods, although the conditional posteriors of δ_{ij}, Ψ_i and $\Sigma_{\varepsilon i}$ are known. This implies that we can employ a simplified version of the Gibbs sampler outlined in George *et al.* (2008), where we start drawing Ψ_i from its full conditional posterior, which follows a Normal distribution. In the next step, we draw the latent variable δ_{ij} from a Bernoulli distribution and in the last step we draw $\Sigma_{\varepsilon i}$ from an inverse Wishart distribution.[5] This algorithm is repeated n times and the first n_{burn} draws are discarded as burn-ins. Averaging the draws of δ_{ij} leads to posterior inclusion probabilities for each variable j. Further details are provided in Appendix D.

[5]In contrast to the implementation in George *et al.* (2008), we impose an inverse Wishart prior on $\Sigma_{\varepsilon i}$ and depart from using a restriction search over $\Sigma_{\varepsilon i}$.

4 Empirical Results

4.1 Data and Model Specification

The bulk of the empirical literature employing GVARs use the dataset put forward in Dees *et al.* (2007b,a), which covers the most important economies in terms of real activity. For this dataset, time series are available from the early 1980s onward. Other studies have extended the country coverage to feature more emerging economies, at the price of limiting the available time span (Feldkircher, 2015; Feldkircher & Huber, 2015). In this paper, our aim is to reserve a significant share of our available time span for forecast evaluation, which is why we opt to have rather long time series – at the implied cost of reducing the country coverage.[6]

We rely on data provided in Dovern *et al.* (2015), that extend the dataset used in Dees *et al.* (2007a,b) with respect to variable coverage and time span. In what follows, we use quarterly data for 36 countries spanning the period from 1979:Q2 to 2013:Q4.[7]

The country-specific VARX* models include seven domestic variables. Six variables are the same as in Dees *et al.* (2007a,b) and Pesaran *et al.* (2009), namely real GDP (y), the change of the (log) consumer price level (Δp), real equity prices (eq), the real exchange rate (e) vis-á-vis the US dollar, and short-term (i_s) and long-term interest rates (i_l). We enlarge this set of variables to feature total credit (tc, domestic and cross-border credit), as a seventh variable. This seems to be important, as the hold-out sample for our forecasting exercise contains the global financial crisis, which spread via both the trade and the financial channel.[8] Note that not all variables are available for each of the countries we consider in this study. With the exception of long-term interest rates, the cross-country coverage of all variables is, however, above 80%. Long-term interest rate data are missing for emerging markets that are characterized by underdeveloped capital markets.

The vector of domestic variables for a typical country i is thus given by

$$\boldsymbol{x}_{it} = (y_{it}, \Delta p_{it}, e_{it}, eq_{it}, i_{sit}, i_{lit}, tc_{it})'. \tag{4.1}$$

[6]Note that this is in contrast to the working paper version of this study available at `http://www.oenb.at/Publikationen/Volkswirtschaft/Working-Papers/2014/Working-Paper-189.html`, which also includes a forecast comparison to the standard, cointegrated GVAR model put forward in Pesaran *et al.* (2004) and Dees *et al.* (2007a,b), among others.

[7]The following countries are included in the respective regions: Europe includes Austria (AT), Belgium (BE), Germany (DE), Spain (ES), Finland (FI), France (FR), Greece (GR), Italy (IT), Netherlands (ND), Portugal (PT), Denmark (DK), Great Britain (GB), Switzerland (CH), Norway (NO) and Sweden (SE). Other Developed economies feature Australia (AU), Canada (CA), Japan (JP), New Zealand (NZ) and the US (US). Emerging Asia includes China (CN), India (IN), Indonesia (ID), Malaysia (MY), Korea (KR), Philippines (PH), Singapore (SG) and Thailand (TH). Latin America comprises Argentina (AR), Brazil (BR), Chile (CL), Mexico (MX) and Peru (PE). Mid-East and Africa consists of Turkey (TR), Saudi Arabia (SA) and South Africa (ZA).

[8]For a more detailed description, see Table A.1 in Appendix A.

We follow the bulk of the literature by including oil prices ($poil_t$) as a global control variable. With the exception of the bilateral real exchange rate, we construct foreign counterparts for all domestic variables. The weights to calculate foreign variables are based on average bilateral annual trade flows in the period from 1980 to 2003, which denotes the end of our initial estimation sample.[9] For a typical country i the set of weakly exogenous and global control variables comprises the following variables,

$$\boldsymbol{x}_{it}^{*} = (y_{it}^{*}, \Delta p_{it}^{*}, eq_{it}^{*}, i_{s\,it}^{*}, i_{l\,it}^{*}, tc_{it}^{*}, poil_{t}^{*})'. \tag{4.2}$$

The US model ($i = 0$) deviates from the other country specifications in that the oil price ($poil_t$) is determined within that country model and the trade weighted real exchange rate (e_t^*) is included to control for co-movements of currencies,

$$\boldsymbol{x}_{0t} = (y_{0t}, \Delta p_{0t}, eq_{0t}, i_{s0t}, i_{l0t}, tc_{0t}, poil_t)', \tag{4.3}$$

$$\boldsymbol{x}_{0t}^{*} = (y_{0t}^{*}, \Delta p_{0t}^{*}, e_{0t}^{*}, eq_{0t}^{*}, i_{s0t}^{*}, i_{l0t}^{*}, tc_{0t}^{*})'. \tag{4.4}$$

The dominant role of the US economy for global financial markets is often accounted for by including a limited set of weakly exogenous variables in its country-specific model. For all countries considered, we set the lag length of endogenous and weakly exogenous variables equal to five. Given the quarterly frequency of the data and the fact that we introduce Bayesian shrinkage, this seems to be a reasonable choice. We correct for outliers in countries that witnessed extraordinarily strong crisis-induced movements in some of the variables contained in our data. We opted to smooth the relevant time series in these cases rather than include step dummies. While step dummies might control for outliers within the specific country model, extreme shocks might still be carried over to other country models via trade-weighted foreign variables. Obviously, this is not the case when smoothing the series in the first place.[10]

[9] Note that recent contributions (Eickmeier & Ng, 2015; Dovern & van Roye, 2014) suggest using financial data to compute foreign variables related to the financial side of the economy (e.g., interest rates or credit volumes). Since our data sample starts in the early 1980s, reliable data on financial flows – such as portfolio flows or foreign direct investment – are not available. See Feldkircher & Huber (2015) for a sensitivity analysis with respect to the choice of weights.

[10] We define outliers as those observations that exceed 1.5 times the interquartile range in absolute value. The identified outliers are then smoothed using cubic spline interpolation techniques and in case they are located at the beginning or the end of the sample – extrapolation techniques. Using the definition of the interquartile range, we identify 2% of our sample as unusual observations. From these 2%, about 60% regard unusual observations for inflation at the beginning of the observation sample. Short-term interest rates (20%) and the real exchange rate (13%) have historically also shown very volatile patterns for the countries covered in this study. More detailed country-specific information is available from the authors upon request.

4.2 Selection of Hyperparameters

Due to the strong heterogeneity observed in the world economy, it is daunting to assume that different countries obey the same structural dynamics in terms of macroeconomic fundamentals. Thus, using the same set of hyperparameters when eliciting the prior for all countries considered might be too restrictive to unveil differences between economies.

The conjugate priors rely on a set of (presumably fixed) hyperparameters which are homogeneous across countries. A specific set of hyperparameters could, however, induce a tight prior in one country while being relatively loose in other countries. To avoid this problem, we follow Carriero *et al.* (2015) and choose the hyperparameters by maximizing the marginal likelihood on a discrete grid of values for $\underline{\alpha}_1$ and $\underline{\alpha}_2$. For the remaining parameters, we set $\underline{\alpha}_{i3} = 100$ and $\theta_{i1} = \theta_{i2} = 1$. For the natural conjugate prior, the marginal likelihood is available in closed form (see, for instance, Bauwens *et al.*, 2000; Koop, 2013). Using the marginal likelihood as a loss function is motivated by the fact that it can be written as a sequence of one-step-ahead predictive densities. Thus maximizing the marginal likelihood under a flat prior is equivalent to minimizing the one-step-ahead prediction errors (Geweke, 2001; Geweke & Whiteman, 2006).[11] Since the marginal likelihood is not available in closed form for the standard Minnesota prior, we use the hyperparameters obtained from the conjugate prior for this setting. Following Carriero *et al.* (2015), the parameter controlling the degree of shrinkage on "other" variables is set equal to 0.8 . Finally, we set $\underline{a}_{ij}$ equal to unity for the first lag of "own" variables.

For the SSVS prior, we set the prior inclusion probability for each variable equal to 0.5, which implies that a priori, every variable is assumed equally likely to enter the model. We set $\tau_{i,0j} = 0.1 s_{ij}$ and $\tau_{i,1j} = 10 s_{ij}$ and rely on the semi-automatic approach described in George *et al.* (2008) to scale the hyperparameters, where s_{ij} is the standard error attached to coefficient j based on a VARX* estimated by OLS in country i. Finally, we set $\underline{S}_i = 10 I_{k_i}$, the prior degrees of freedom $\underline{v}_i$ to k_i and $\underline{\Psi}_i$ equal to the zero matrix. As a robustness check we also used a standard random walk prior specification in combination with the SSVS prior. This implies that the prior mean on the first own lag is set equal to unity. However, since almost no shrinkage is imposed for the case when $\delta_{ij} = 1$ the results are rather similar with the standard SSVS implementation. Thus for the sake of brevity we only report the results obtained by setting the prior mean equal to zero for all coefficients. To assess the importance of shrinkage we include in the forecast exercise a prior that is flat over the coefficients (diffuse). This prior is implemented by setting $\underline{\alpha}_{i1} = \underline{\alpha}_{i2} = \underline{\alpha}_{i3} = 10^{10}$ in the conjugate Minnesota prior setup.

[11] Note that this differs from the procedure proposed by Giannone *et al.* (2012), since we do not integrate out the hyperparameters in a Bayesian fashion but simply plug in an estimate of the posterior mode of $\underline{\alpha}_1$ and $\underline{\alpha}_2$ under a diffuse prior. This approach seems convenient since it avoids MCMC sampling for the conjugate priors, which proves to be important for the empirical application.

4.3 Forecast Performance

The initial estimation period ranges from 1979:Q2 to 2003:Q4 and we use the period 2004:Q1-2013:Q4 as out-of-sample hold-out observations to compare predictive performance across specifications. We base our comparison on recursive one-quarter-ahead and four-quarters-ahead predictions obtained by re-estimating the models over an expanding window defined by the beginning of the available sample and the corresponding period in the hold-out sample.[12] In what follows, we compare the forecast performance for both point and density forecasts. For means of comparison we choose the root mean squared error (RMSE) for point forecasts and log predictive scores (LPS) to evaluate density forecasts.[13] All forecasts are benchmarked to those of a fifth-order univariate autoregressive (AR(5)) model and forecast errors are reported in an unweighted fashion. That is, we do not attach more weight to favor GVAR specifications that improve forecasts for particular countries which stand out in terms of economic activity. The AR(5) model is estimated in a Bayesian fashion, where a shrinkage prior in the spirit of a standard Minnesota prior is imposed (i.e., higher lag orders are strongly pushed towards zero) and the hyperparameter is estimated by maximizing the ML over the same grid as given above.

Forecast Evaluation: Overall Results

Overall results based on one-step-ahead forecasts are provided in the upper panel of Table 1, while those for four-steps-ahead are presented in the bottom panel. We start by considering the best performing settings in the aggregate, across all variables and both prediction horizons. To assess which specification exhibits the best overall performance, the final row in both panels of Table 1 depicts the average RMSE across variables and the sum of variable-specific LPS.[14] Since the LPS is a scoring rule that takes into account higher order moments of the corresponding one-step and four-steps-ahead predictive densities, we evaluate the overall predictive fit of a given model by means of the LPS. For both time horizons considered, the B-GVAR coupled with the SSVS prior outperforms all competing specifications. Models that use the conjugate priors (M-C and S&Z) also tend to exhibit a compelling overall performance, performing slightly worse. Note that the Minnesota prior with a fixed covariance matrix of the error term (M-NC) shows the weakest performance among all shrinkage priors considered.[15] Basing our conclusion on the overall sum of variable-specific LPS values, the best overall

[12] This implies that the marginal likelihood is numerically optimized for each time point in our hold-out sample.

[13] See Appendix E for more details on the construction of the density measures.

[14] We use the sum of variable-specific LPS due to numerical reasons associated with the evaluation of a high-dimensional predictive density.

[15] In order to differentiate this prior from the standard conjugate Minnesota prior, that integrates out uncertainty with respect to $\Sigma_{\varepsilon i}$, we include it under the header "Non-Conjugate" in Table 1, although (conditional) conjugacy would be present in this setting.

predictive performance is thus achieved by the Bayesian GVAR model with an SSVS prior. In order to isolate the improvements in out-of-sample predictive accuracy that emanate from the explicit assessment of international linkages in B-GVAR models, the BVAR specification which is used as a comparison in the results of Table 1 is estimated using this prior.

We start by addressing the results obtained for one-quarter-ahead predictions. Considering results on point forecasts first, most specifications yield forecasts that improve upon the naive model, as indicated by relative RMSEs below unity. Comparing the results of the BVAR specification with that of the GVAR under a diffuse prior does not yield clear-cut results concerning the superiority of one of the two modelling frameworks for all variables considered. Combining the virtues of both approaches (shrinkage and international linkages) boosts forecast performance considerably. All B-GVAR prior settings that induce shrinkage outperform the benchmark model. Only one-step-ahead forecasts for total credit under the two conjugate priors are similar to those of the benchmark in terms of RMSEs. Comparing B-GVAR shrinkage specifications, our results show a particular good forecast performance for the Minnesota prior with fixed variance-covariance matrix for the error term (M-NC). The M-NC prior yields the best point forecasts, outperforming all competitors for six out of the seven variables considered. The forecast performance of the standard conjugate variant of the Minnesota prior (M-C) is less spectacular, yielding prediction accuracy measures close to that of the BVAR under an SSVS prior setting. Finally, the S&Z specification excels in forecasting short-term interest rates, while for the remaining variables forecasts tend to be similar or slightly worse than the ones from the M-NC specification. The SSVS prior also performs well compared to the autoregressive benchmark, achieving a similar aggregate level of predictive accuracy across variables as that of the other B-GVAR alternatives. However, only the B-GVAR with an M-NC prior leads to improvements over the BVAR when comparing average RMSE across variables.

Next, we evaluate the relative quality of density forecasts. Table 1, upper panel right-hand side, displays the sum of log predictive scores over countries per variable, reported as differences to the benchmark autoregressive model. Positive values indicate a better performance of the forecast method under consideration compared to the benchmark. The results confirm some of the findings for point forecast accuracy and provide new insights. Both isolated country-specific VAR models (BVAR) and the international model without shrinkage (diffuse) outperform forecasts of the naive model on average across all variables. Again, there is no clear superiority structure when comparing these two approaches, while forecasts can be further improved by considering models that explicitly feature international linkages coupled with priors that induce shrinkage on the parameters (M-C, S&Z, and SSVS). Overall improvements over forecasts from the benchmark are most pronounced for the SSVS prior and the M-C specification.

The lower panel of Table 1 presents the results based on a four-quarters-ahead forecast horizon. In the medium-term, forecasts from the naive benchmark specification appear hard to beat. The merits from using shrinkage priors play out more strongly in this setting than in the short-term prediction exercise. While forecasts from the isolated country-specific BVAR models do not tend to worsen markedly with the expanded forecast horizon, the B-GVAR model under a diffuse prior setting shows a relatively poor forecasting performance. This holds true for both point and density forecasts. In line with our findings on the one-step-ahead forecast horizon, forecasts can be further improved by considering GVAR specifications coupled with shrinkage priors. However, forecast gains are less pronounced than in the short run. Among the prior specifications considered, the S&Z prior excels in point forecast accuracy, while the SSVS prior yields again the strongest performance regarding density forecasts.

Before providing information on the cross-sectional distribution of forecast accuracy of the competing models, it is worth emphasizing that the forecast performance tends to fluctuate over time. Inspection of the cumulative LPS over the out-of-sample prediction sample reveals that especially within crisis periods, the combination of shrinkage priors coupled with the inclusion of global factors improves the predictive performance significantly.[16] This implies that in circumstances where most economies in our sample experienced a sharp drop in output growth caused by a global shock, information originating from the cross-section becomes increasingly important in terms of improving the predictive ability of econometric specifications.

[Table 1 about here.]

Forecast Evaluation: Cross-Country Differences in Point Forecast Accuracy

In order to examine whether there are systematic cross-country differences in point forecast accuracy for the priors considered, Figures 1 to 5 show the cross-sectional distribution of relative RMSE for the one-step-ahead forecast horizon for different world regions. We present results for Europe, other developed economies, emerging Asia and Latin America.[17] With the exception of Latin America, all plots have the same scaling in order to ease regional comparison of forecast accuracy under the different specifications.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[16] The Figures depicting cumulative LPS over the forecasting period can be found in Appendix F.

[17] We consider all variables but long-term interest rates and equity prices, for which the cross-country coverage is limited. Results for these variables as well as for the four-quarters-ahead forecast horizon are available upon request.

[Figure 4 about here.]

[Figure 5 about here.]

The results indicate that point forecast accuracy varies strongly across regions and less so across variables. Taking a regional stance, the largest dispersion of relative RMSE values is observed for variables in Latin American economies. Here the cross-sectional variance of forecast accuracy tends to be large for practically all prior specifications considered. Point forecasts are particularly inaccurate and cross-sectional distributions wide for inflation, total credit and short-term interest rates. For these variables the distributions are about three to four times larger than the ones for the rest of the regions and medians of the distribution of RMSE tend to indicate a worse performance relative to the naive model. While the dispersion of relative RMSE is markedly smaller in emerging Asia, some of the priors considered yield very inaccurate point forecasts for real activity and inflation in India and Indonesia. By comparison, forecast performance is very homogeneous for variables of European and other developed economies. The distributions of relative predictive ability tend to be very tight, and forecasts that fall far off the median do so for most priors considered. Inaccurate point forecasts for real GDP can be found for Norway, whose economy depends strongly on oil exports, and Denmark. For total credit and the real exchange rate, relative RMSE figures are particularly large in Great Britain and Switzerland, two countries with a large financial sector. Countries that appear as outliers in the box plots might either indicate that some country-specific features (e.g., oil based economy, heavy financial sector) are not correctly reflected in the model, the (linear) specification of the model might be too restrictive or the variance of the underlying time series might be comparably large.

Comparing across prior specifications, these disaggregated results corroborate the findings provided in Table 1. Concentrating on the median value, most prior specifications tend to outperform forecasts from the naive models for all regions and variables with the exception of Latin America. Forecasts under the M-NC and SSVS priors often yield the lowest cross-sectional median of relative RMSEs. At first sight, also the diffuse prior yields quite accurate point forecasts. However, the distribution of relative RMSEs tend to be tighter when using priors that incorporate country-specific shrinkage. This holds in particular true for Latin American variables, where relative RMSE distributions under some shrinkage priors are very tight, while the forecast performance indicator is very disperse under the diffuse prior.

5 Conclusions

In this paper we develop a Bayesian GVAR model and assess its out-of-sample predictive performance in terms of point- and density forecasts. We use a large quarterly dataset that starts in 1979:Q2, excels in country coverage and covers many of the most important macroeconomic and financial variables. This dataset allows us to reserve a significant share of the data as a forecast evaluation sample (40 observations by country, spanning the period from 2004:Q1 to 2013:Q4). Our forecast evaluation sample thus includes periods of very distinct macroeconomic and financial conditions: the period of the great moderation that was accompanied by stable GDP growth and low inflation, followed by the global financial crisis which triggered most of the economies to enter (prolonged) recession phases, and the ongoing period of recovery since then. Evaluating forecasts over that period yields a fair assessment of the usefulness of Bayesian GVAR models, and the length of our hold-out sample significantly improves upon earlier studies on forecasting using GVAR specifications (see, e.g., Pesaran *et al.*, 2009).

Our *main results* are the following. First, we provide ample evidence that taking *international linkages* among the economies into account and using priors that *induce shrinkage on the parameters locally* greatly improves forecast performance. Throughout the set of variables considered in this study, the diffuse prior setup that is flat over the coefficients as well as forecasts from isolated Bayesian VAR models do not tend to rank among the best performing forecast specifications. This holds true for both point forecasts and density forecasts as well as the short-term and medium-term horizon. To set the degree of shrinkage for each country model locally, we numerically optimize the marginal likelihood with respect to the hyperparameters, which is equivalent to minimizing the one-step-ahead prediction errors. This allows us to accommodate a large degree of heterogeneity across the economies, which appears of particular importance for forecasting in a global setting. Second, within the class of Bayesian GVARs that induce country-specific shrinkage, no single prior dominates all forecasting setups. The SSVS prior shows a very strong forecast performance over the hold-out sample, while the forecast improvement of the conjugate Minnesota prior compared to the naive benchmark tends to be small. The Minnesota prior with a non-random variance covariance matrix of the error term and the prior put forward in Sims & Zha (1998) show excellent track records of point forecasts for the one-step-ahead and four-steps-ahead forecast horizons respectively. Moreover, the Minnesota prior with a non-random variance covariance matrix of the error term together with the SSVS prior both excel in short-term density forecasting for the majority of variables considered. The latter also shows a strong density forecast performance in the longer run. *Taking a regional stance*, our results indicate that forecasts for Latin America are particularly inaccurate for most specifications considered in this study. By contrast, forecast performance for variables from European or other developed economies is much more homogeneous and differences between the various prior setups considered more modest.

References

Adolfson, M., J. Lindé, & M. Villani (2007): "Forecasting performance of an open economy dsge model." *Econometric Reviews* **26(2-4)**: pp. 289–328.

Alessi, L. & M. Bańbura (2009): "Comparing global macroeconomic forecasts." Mimeo.

Bańbura, M., D. Giannone, & L. Reichlin (2010): "Large bayesian vector auto regressions." *Journal of Applied Econometrics* **25(1)**: pp. 71–92.

Bauwens, L., M. Lubrano, & J.-F. Richard (2000): *Bayesian inference in dynamic econometric models*. Oxford University Press.

Carriero, A., T. E. Clark, & M. Marcellino (2015): "Bayesian VARs: Specification Choices and Forecast Accuracy." *Journal of Applied Econometrics* **30(1)**: pp. 46–73.

Chudik, A. & M. Fratzscher (2011): "Identifying the global transmission of the 2007-2009 financial crisis in a GVAR model." *European Economic Review* **55(3)**: pp. 325–339.

Dees, S., S. Holly, H. M. Pesaran, & V. L. Smith (2007a): "Long Run Macroeconomic Relations in the Global Economy." *Economics - The Open-Access, Open-Assessment E-Journal* **1(3)**: pp. 1–20.

Dees, S., F. di Mauro, H. M. Pesaran, & L. V. Smith (2007b): "Exploring the international linkages of the euro area: a global VAR analysis." *Journal of Applied Econometrics* **22(1)**.

Doan, T. R., B. R. Litterman, & C. A. Sims (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* **3**: pp. 1–100.

Dovern, J., M. Feldkircher, & F. Huber (2015): "Does Joint Modeling of the World Economy Pay Off? Evaluating Multivariate Forecasts from a Bayesian GVAR." *Working Paper Series 200/2015*, Oesterreichische Nationalbank.

Dovern, J. & B. van Roye (2014): "International transmission and business-cycle effects of financial stress." *Journal of Financial Stability* **13(0)**: pp. 1 – 17.

Eickmeier, S. & T. Ng (2015): "How do us credit supply shocks propagate internationally? a gvar approach." *European Economic Review* **74(0)**: pp. 128 – 145.

Feldkircher, M. (2015): "A Global Macro Model for Emerging Europe." *Journal of Comparative Economics* **forthcoming**.

Feldkircher, M. & F. Huber (2015): "The international transmission of US shocks – Evidence from Bayesian global vector autoregressions." *European Economic Review* **forthcoming**.

Galesi, A. & S. Sgherri (2013): *The GVAR Handbook: Structure and Applications of a Macro Model of the Global Economy for Policy Analysis*, chapter Regional financial spillovers across Europe, pp. 255–270. Oxford University Press.

Garrat, A., K. Lee, M. H. Pesaran, & Y. Shin (2006): *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*. Oxford University Press.

George, E. I. & R. McCulloch (1993): "Variable selection via Gibbs sampling." *Journal of the American Statistical Association* **88**: pp. 881–889.

George, E. I., D. Sun, & S. Ni (2008): "Bayesian stochastic search for VAR model restrictions." *Journal of Econometrics* **142(1)**: pp. 553–580.

Geweke, J. (2001): "Bayesian econometrics and forecasting." *Journal of Econometrics* **100(1)**: pp. 11–15.

Geweke, J. & G. Amisano (2010): "Comparing and evaluating bayesian predictive distributions of asset returns." *International Journal of Forecasting* **26(2)**: pp. 216–230.

Geweke, J. & C. Whiteman (2006): "Bayesian forecasting." *Handbook of economic forecasting* **1**: pp. 3–80.

Giannone, D., M. Lenza, D. Momferatou, & L. Onorante (2013): "Short-term inflation projections: a Bayesian vector autoregressive approach." *International Journal of Forecasting* .

Giannone, D., M. Lenza, & G. E. Primiceri (2012): "Prior selection for vector autoregressions." *Technical report*, National Bureau of Economic Research.

Giannone, D. & L. Reichlin (2009): "Comments on "Forecasting economic and financial variables with global VARs"." *International Journal of Forecasting* **25(4)**: pp. 684–686.

Greenwood-Nimmo, M., V. H. Nguyen, & Y. Shin (2012): "Probabilistic Forecasting of Output Growth, Inflation and the Balance of Trade in a GVAR Framework." *Journal of Applied Econometrics* **27**: pp. 554–573.

Han, F. & T. H. Ng (2011): "ASEAN-5 Macroeconomic Forecasting Using a GVAR Model." *Asian Development Bank, Working Series on Regional Economic Integration* **76**.

Kadiyala, K. & S. Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics* **12(2)**: pp. 99–132.

Karlsson, S. (2012): "Forecasting with Bayesian Vector Autoregressions." *Örebro University Working Paper* **12/2012**.

Koop, G. & D. Korobilis (2015): "Model uncertainty in panel vector autoregressive models." *European Economic Review* **forthcoming**.

Koop, G. M. (2013): "Forecasting with medium and large bayesian vars." *Journal of Applied Econometrics* **28(2)**: pp. 177–203.

Korobilis, D. (2015): "Prior selection for panel vector autoregressions." *MPRA Paper 64143*, University Library of Munich.

Litterman, R. (1986): "Forecasting with Bayesian vector autoregressions - Five years of experience." *Journal of Business and Economic Statistics* **5**: pp. 25–38.

Pesaran, M. H., T. Schuermann, & L. V. Smith (2009): "Forecasting economic and financial variables with global VARs." *International Journal of Forecasting* **25(4)**: pp. 642–675.

Pesaran, M. H., T. Schuermann, & S. M. Weiner (2004): "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model." *Journal of Business and Economic Statistics, American Statistical Association* **22**: pp. 129–162.

Rodgers, D. P. (1985): "Improvements in multiprocessor system design." In "ACM SIGARCH Computer Architecture News," volume 13, pp. 225–231. IEEE Computer Society Press.

Sims, C. A. (1992): "Bayesian Inference for Multivariate Time Series with Trend." *presented at the American statistical Association meetings* .

Sims, C. A. & T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* **39(4)**: pp. 949–968.

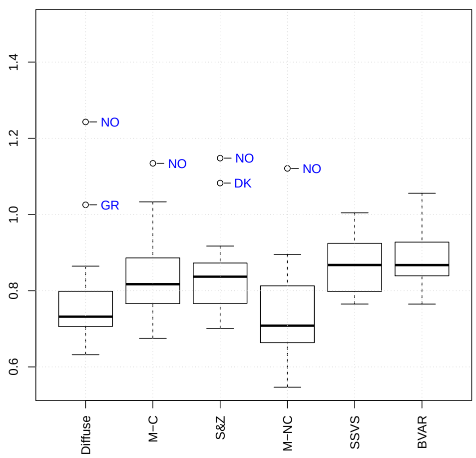
Table 1: Forecast performance relative to autoregressive benchmark

	RMSE						LPS					
	One-quarter-ahead											
		Conjugate		Non-Conjugate				Conjugate		Non-Conjugate		
	Diffuse	M-C	S&Z	M-NC	SSVS	BVAR	Diffuse	M-C	S&Z	M-NC	SSVS	BVAR
y	0.84	0.90	0.92	**0.79**	0.90	0.91	130.70	175.55	166.49	132.16	**206.60**	180.66
Δp	0.99	**0.90**	**0.90**	**0.90**	0.99	0.92	10.13	12.62	13.07	**15.85**	1.30	1.53
e	0.96	0.99	0.99	**0.93**	0.95	0.96	24.90	30.27	**33.20**	24.78	17.25	9.45
i_S	1.07	0.86	**0.82**	0.84	0.85	0.85	8.87	3.82	4.27	**11.79**	3.59	2.43
i_L	0.82	0.84	0.84	**0.75**	0.87	0.89	5.01	6.42	5.09	**8.62**	-3.06	-3.11
eq	0.94	0.86	0.88	**0.75**	0.85	0.85	274.58	276.40	275.32	279.71	**285.53**	241.60
tc	1.03	1.01	1.00	**0.92**	0.93	0.94	57.68	204.71	201.50	128.56	**282.16**	262.14
$\varnothing/\sum$	0.95	0.91	0.91	**0.84**	0.91	0.90	511.87	709.79	698.94	601.47	**793.37**	694.69
	Four-quarters-ahead											
		Conjugate		Non-Conjugate				Conjugate		Non-Conjugate		
	Diffuse	M-C	S&Z	M-NC	SSVS	BVAR	Diffuse	M-C	S&Z	M-NC	SSVS	BVAR
y	1.81	0.94	**0.88**	1.22	1.03	0.97	102.04	149.38	147.14	109.77	**163.10**	145.57
Δp	1.83	1.02	**0.97**	1.49	1.11	0.98	-12.92	1.00	**3.03**	-2.45	-11.32	-13.27
e	2.13	1.18	1.07	1.74	1.04	**1.02**	-6.27	11.90	**13.58**	4.94	10.97	6.10
i_S	2.07	1.03	**0.93**	1.30	1.05	0.97	-1.49	2.47	3.22	**8.06**	4.61	0.40
i_L	1.86	1.01	**0.96**	1.17	0.97	1.03	-5.04	3.75	3.09	**4.96**	-4.93	-7.82
eq	2.93	1.09	1.04	1.92	1.26	**0.92**	128.57	145.43	147.12	133.33	**148.07**	125.81
tc	1.22	0.72	0.68	0.90	**0.67**	0.69	115.32	204.63	201.33	150.82	**240.94**	227.55
$\varnothing/\sum$	1.98	1.00	**0.94**	1.39	1.02	0.95	320.21	518.57	518.52	409.42	**551.43**	484.33

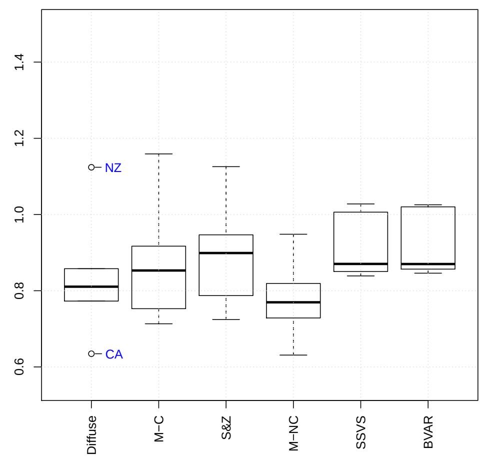
Notes: The figures on the left panel refer to root mean squared error (RMSE) relative to an autoregressive model of order five (AR(5)). Values below one indicate better performance of the respective forecast model compared to the benchmark. Figures on the right panel refer to the differences of the sum of variable specific log predictive scores (LPS) relative to the AR(5) benchmark. The rows starting with $\varnothing/\sum$ depict the average RMSE and the sum of the variable-specific LPS for all variables considered. Results based on forecasts over the time period 2004:Q1-2013:Q4. Figures for best performing models are in bold. *Diffuse* stands for the model estimated using maximum likelihood, *M-C* denotes the GVAR with the conjugate variant of the Minnesota prior, *S&Z* refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), *M-NC* stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, *SSVS* denotes the GVAR estimated using the SSVS prior and *BVAR* denotes a set of isolated, country-specific vector autoregressions estimated using the SSVS prior.

Figure 1: Cross-sectional distribution of 1-step-ahead RMSE values for real GDP

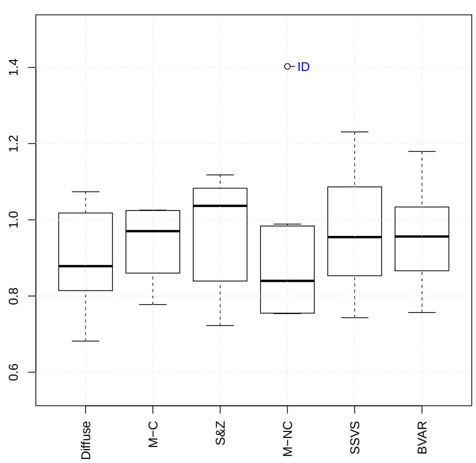
(a) Europe



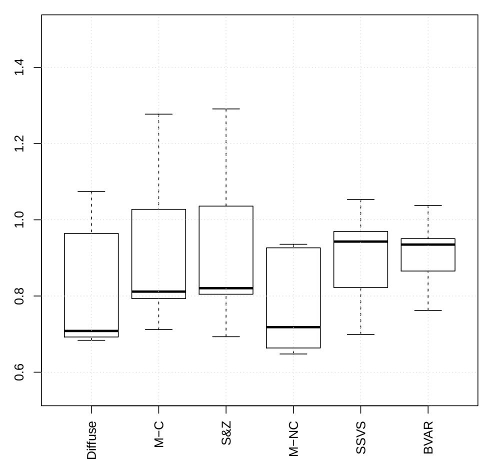
(b) Other developed economies



(c) Emerging Asia



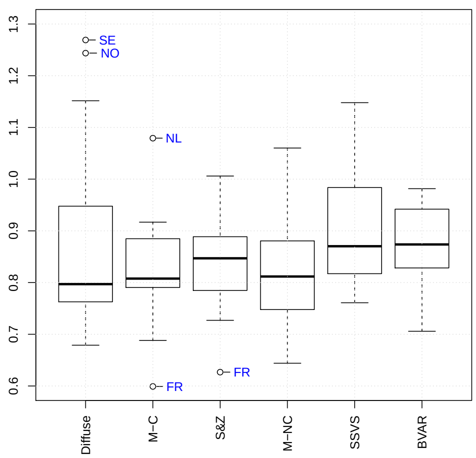
(d) Latin America



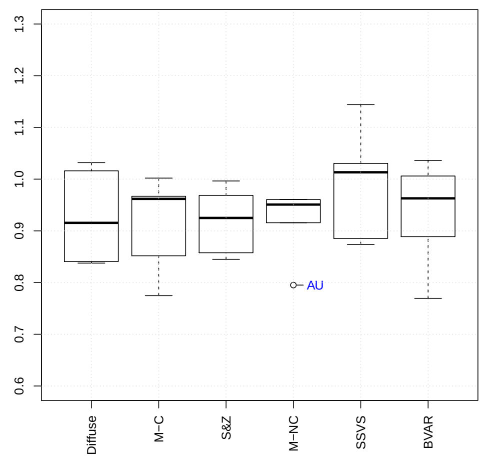
Notes: The figures show the cross-sectional distribution of the ratio of the RMSE corresponding to the model to the RMSE of an autoregressive model of order five over the time period 2004:Q1-2013:Q4. Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed. Observations that exceed 1.5 times the interquartile range are marked as outliers.

Figure 2: Cross-sectional distribution of 1-step-ahead RMSE values for inflation

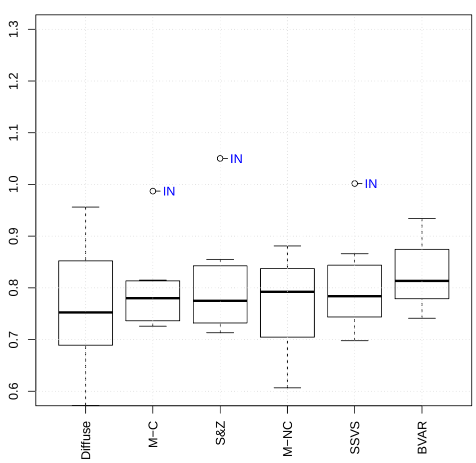
(a) Europe



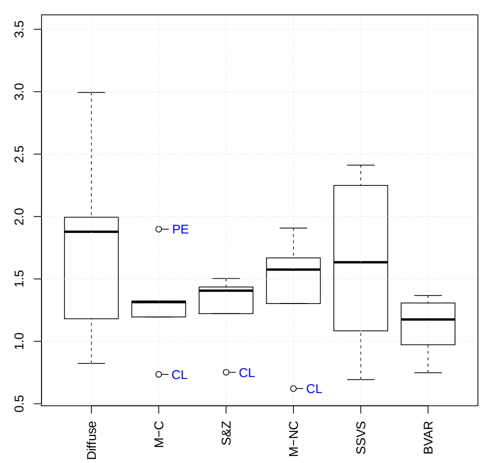
(b) Other developed economies



(c) Emerging Asia



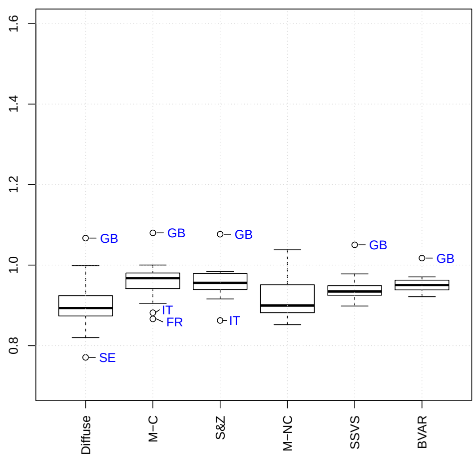
(d) Latin America



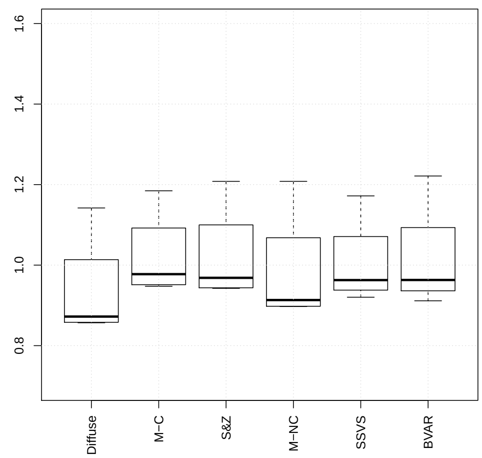
Notes: The figures show the cross-sectional distribution of the ratio of the RMSE corresponding to the model to the RMSE of an autoregressive model of order five over the time period 2004:Q1-2013:Q4. Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed. Observations that exceed 1.5 times the interquartile range are marked as outliers.

Figure 3: Cross-sectional distribution of 1-step-ahead RMSE values for the real exchange rate

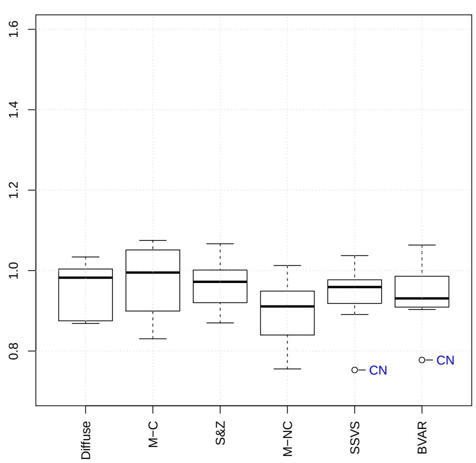
(a) Europe



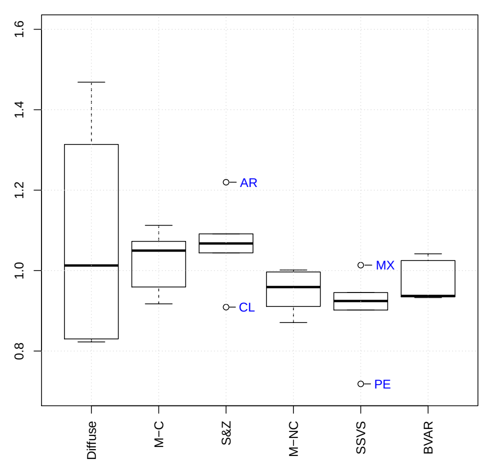
(b) Other developed economies



(c) Emerging Asia



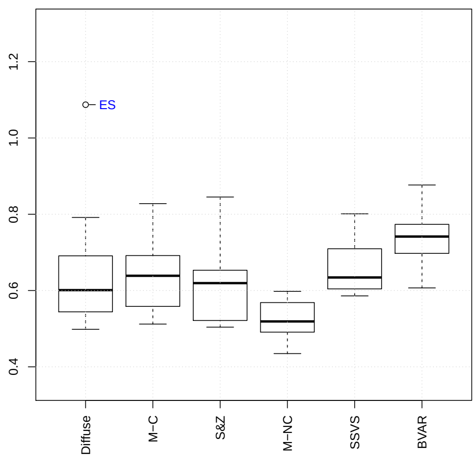
(d) Latin America



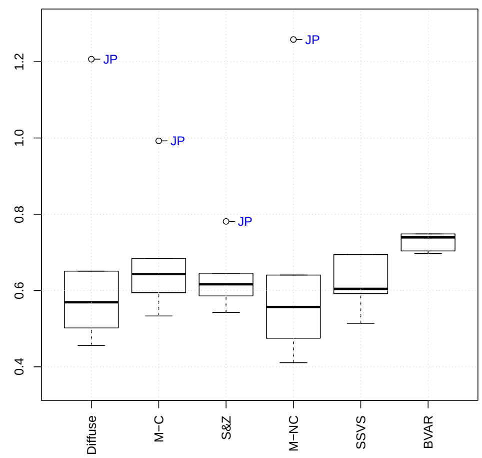
Notes: The figures show the cross-sectional distribution of the ratio of the RMSE corresponding to the model to the RMSE of an autoregressive model of order five over the time period 2004:Q1-2013:Q4. Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed. Observations that exceed 1.5 times the interquartile range are marked as outliers.

Figure 4: Cross-sectional distribution of 1-step ahead RMSE values for short-term interest rates

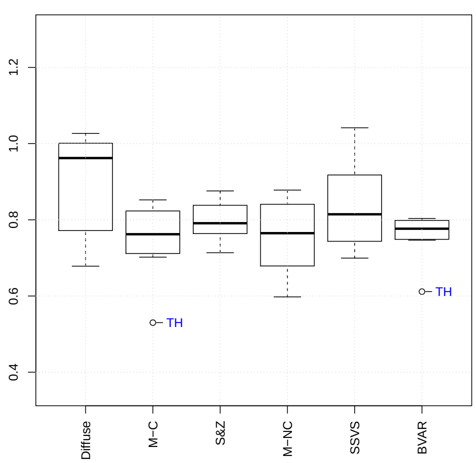
(a) Europe



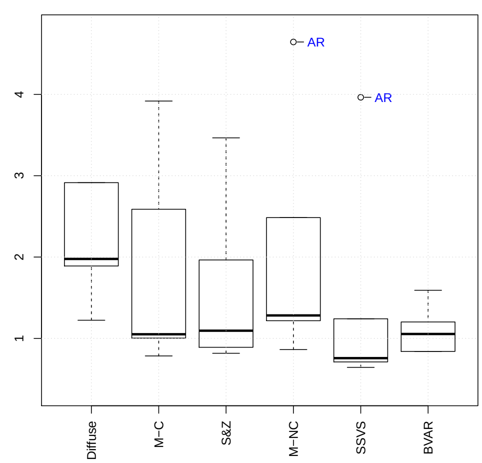
(b) Other developed economies



(c) Emerging Asia



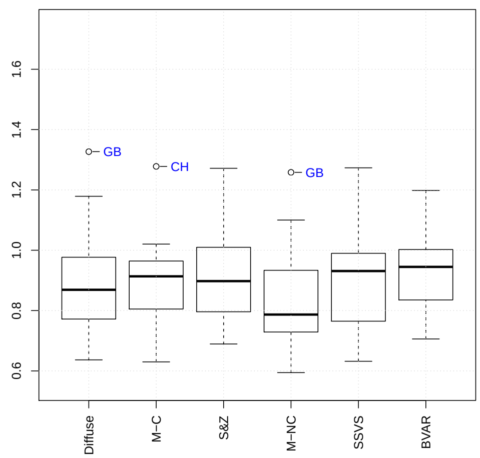
(d) Latin America



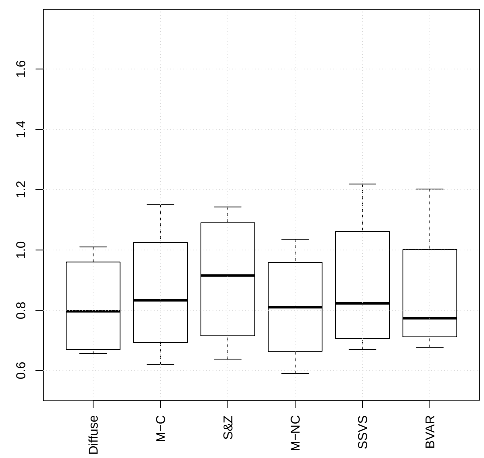
Notes: The figures show the cross-sectional distribution of the ratio of the RMSE corresponding to the model to the RMSE of an autoregressive model of order five over the time period 2004:Q1-2013:Q4. Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed. Observations that exceed 1.5 times the interquartile range are marked as outliers.

Figure 5: Cross-sectional distribution of 1-step-ahead RMSE values for total credit

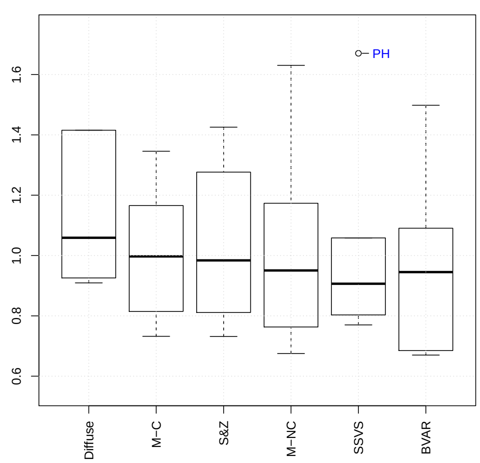
(a) Europe



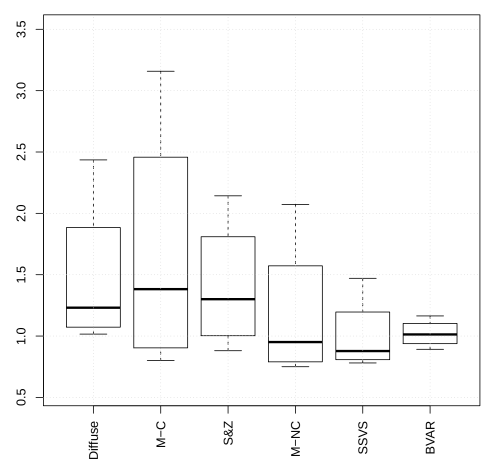
(b) Other developed economies



(c) Emerging Asia



(d) Latin America



Notes: The figures show the cross-sectional distribution of the ratio of the RMSE corresponding to the model to the RMSE of an autoregressive model of order five over the time period 2004:Q1-2013:Q4. Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed. Observations that exceed 1.5 times the interquartile range are marked as outliers.

Web Appendix to “Forecasting with Global Vector Autoregressive Models: A Bayesian Approach” by J. Crespo Cuaresma, M. Feldkircher and F. Huber

Appendix A Data Description

Table A.1: Data description

Variable	Description	Min.	Mean	Max.	Coverage
y	Real GDP, average of 2005=100. Seasonally adjusted, in logarithms.	2.173	4.298	5.400	100%
Δp	Consumer price inflation. CPI seasonally adjusted, in logarithms.	-0.157	0.021	0.660	100%
e	Nominal exchange rate vis-à-vis the US dollar, deflated by national price levels (CPI).	-5.373	-2.814	4.968	97.2%
i_S	Typically 3-months-market rates, rates per annum.	-0.001	0.118	5.189	97.2%
i_L	Typically government bond yields, rates per annum.	0.000	0.077	0.275	61.1%
tc	Total credit (domestic + cross border), seasonally adjusted, in logarithms, average of 2005=100.	-14.140	3.514	6.552	83.33%
$poil$	Price of oil, seasonally adjusted, in logarithms.	-	-	-	-
Trade flows	Bilateral data on exports and imports of goods and services, annual data.	-	-	-	-

Notes: Summary statistics pooled over countries and time. The coverage refers to the cross-country availability per country, in %. Data are from the IMF's IFS data base and national sources. Trade flows stem from the IMF's DOTS data base. For more details see the data appendix in Feldkircher (2015).

Appendix B Deriving the GVAR Model

For the sake of exposition, let us assume that $p = 1$, $p^* = 1$ and $a_{i0} = 0$. Following Pesaran *et al.* (2004), the country-specific models in equation (2.1) can be rewritten as

$$A_i z_{it} = B_i z_{it-1} + \varepsilon_{it}, \tag{B.1}$$

where $A_i = (I_{k_i}, -\Lambda_{i0})$, $B_i = (\Phi_i, -\Lambda_{i1})$ and $z_{it} = (x'_{it}, x^{*\prime}_{it})'$. By defining a suitable link matrix W_i of dimension $(k_i + k_i^*) \times k$, where $k = \sum_{i=0}^{N} k_i$, we can rewrite z_{it} as $z_{it} = W_i x_t$, with x_t (the so-called global vector) being a vector where all the endogenous variables of the countries in our sample are stacked, i.e., $x_t = (x'_{0t}, \ldots, x'_{Nt})'$. Replacing z_{it} with $W_i x_t$ in (B.1) and stacking the different local models leads yields the global model,

$$\begin{aligned} x_t &= G^{-1} H x_{t-1} + G^{-1} \epsilon_t \\ &= F x_{t-1} + e_t. \end{aligned} \tag{B.2}$$

Here, $G = ((A_0 W_0)', \cdots, (A_N W_N)')'$ and $H = ((B_0 W_0)', \cdots, (B_N W_N)')'$ denote the corresponding stacked matrices containing the parameter matrices of the country-specific specifications. In line with existing work (e.g., Dees *et al.*, 2007b) we assume that G is invertible. Finally, $e_t \sim \mathcal{N}(0, \Sigma_e)$, where $\Sigma_e = G^{-1} \Sigma_\epsilon (G^{-1})'$ and Σ_ϵ is a block-diagonal matrix given by

$$\Sigma_\epsilon = \begin{pmatrix} \Sigma_{\varepsilon 0} & 0 & \cdots & 0 \\ 0 & \Sigma_{\varepsilon 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{\varepsilon N} \end{pmatrix}. \tag{B.3}$$

Consequently, the matrix G establishes contemporaneous cross-country correlations. The eigenvalues of the matrix F provide information about the stability of the global system. In the empirical application we rule out explosive behavior of the model by discarding posterior draws that significantly fall outside the unit circle.[18] The framework outlined above deviates from the work pioneered by Pesaran *et al.* (2004) in that we do not explicitly impose cointegration relationships in the individual country-specific models.

[18] The proportion of such draws is extremely small and including them in the analysis does not qualitatively affect any of the conclusions of the study.

Appendix C Posterior Distributions

The Conjugate Case

For all priors discussed in Section 3 that can be cast into a form that uses dummy observations, prior quantities can be expressed as

$$\underline{\Pi}_i = (\underline{Z}_i'\underline{Z}_i)^{-1}\underline{Z}_i'\underline{x}_i \tag{C.1}$$

$$\underline{V}_i = (\underline{Z}_i'\underline{Z}_i)^{-1} \tag{C.2}$$

$$\underline{S}_i = (\underline{x}_i - \underline{Z}_i\underline{\Pi}_i)'(\underline{x}_i - \underline{Z}_i\underline{\Pi}_i) \tag{C.3}$$

where $\underline{Z}_i, \underline{x}_i$ denotes any (or a combination) of the dummy observations discussed in Section 3.

In the conjugate case, the posterior distributions of Ψ_i and $\Sigma_{\varepsilon i}$ are of Normal and inverse-Wishart form, respectively. Formally, this implies that

$$\Psi_i | \Sigma_{\varepsilon i}, \mathcal{D}_T \sim \mathcal{N}(\overline{\Psi}_i, \Sigma_{\varepsilon i} \otimes \overline{V}_i) \tag{C.4}$$

$$\Sigma_{\varepsilon i} | \Psi_i, \mathcal{D}_T \sim \mathcal{IW}(\overline{S}_i, \overline{v}_i) \tag{C.5}$$

where $\mathcal{D}_T$ denotes the data up to time T. The posterior mean of $\Psi_i = \text{vec}(\Pi_i)$ is given by

$$\overline{\Pi}_i = (\overline{Z}_i'\overline{Z}_i)^{-1}\overline{Z}_i'\overline{x}_i, \tag{C.6}$$

where $\overline{Z}_i$ and $\overline{x}_i$ denote the dummy-observation-augmented data matrices. Moreover, $\overline{V}_i$ is simply

$$\overline{V}_i = (\overline{Z}_i'\overline{Z}_i)^{-1} \tag{C.7}$$

The scale matrix of the posterior of $\Sigma_{\varepsilon i}$ is given by

$$\overline{S}_i = (\overline{x}_i - \overline{Z}_i\overline{\Pi}_i)'(\overline{x}_i - \overline{Z}_i\overline{\Pi}_i) \tag{C.8}$$

and the posterior degrees of freedom are $T + v_i$. The conjugate nature of this prior implies that posterior distributions are available in closed-form.

The Non-Conjugate / SSVS Case

Following George *et al.* (2008), we replace $\Sigma_{\varepsilon i} \otimes \overline{V}_i$ in equation (C.4) by a $v_i \times v_i$ matrix $\overline{R}_i$, where

$$\overline{R}_i = \left(\Sigma_{\varepsilon i}^{-1} \otimes (Z_i'Z_i) + \underline{R}_i^{-1}\right)^{-1}. \tag{C.9}$$

The mean of the conditional posterior is given by

$$\overline{\Psi}_i = \overline{R}_i \left(\underline{R}_i^{-1}\underline{\Psi}_i + \Sigma_{\varepsilon i}^{-1} \otimes (Z_i'x_i')\right). \tag{C.10}$$

The posterior degrees of freedom are still $\overline{v}_i = T + \underline{v}_i$ and the posterior scale matrix is given by

$$\overline{S}_i = \underline{S}_i + (x_i - Z_i\overline{\Pi}_i)'(x_i - Z_i\overline{\Pi}_i). \tag{C.11}$$

Finally, the conditional posterior of δ_{ij} is distributed as Bernoulli,

$$\delta_{ij}|\delta_{i\bullet}, \Psi_i, \Sigma_{\varepsilon i}, \mathcal{D}_{iT} \sim \text{Bernoulli}(\overline{q}_{ij}) \tag{C.12}$$

where the notation $\delta_{i\bullet}$ indicates conditioning on all δ_{ig} for $g \neq j$ and the probability that $\delta_{ij} = 1$ is given by

$$\overline{q}_{ij} = \frac{\frac{1}{\tau_{1j}} \exp(-\frac{\Psi_{ij}^2}{2\tau_{1j}})}{\frac{1}{\tau_{1j}} \exp(-\frac{\Psi_{ij}^2}{2\tau_{1j}})\underline{q}_{ij} + \frac{1}{\tau_{0j}} \exp(-\frac{\Psi_{ij}^2}{2\tau_{0j}})(1 - \underline{q}_{ij})}. \tag{C.13}$$

Appendix D Posterior Inference at the Global Level: The Implications and Advantages of Country-Specific Priors

The method described in Section 3 imposes priors exclusively at the individual country level. The main reason for local prior elicitation is computational. Furthermore, it is straightforward to show that placing the priors locally leads to the same priors on the global level scaled by the strength of the invoked trade links.

Prior implications at the global level

The global implications of a prior imposed locally and the corresponding prior variances can be derived by substituting $W_i x_t$ in equation (B.1),

$$A_i W_i x_t = B_i W_i x_{t-1} + \varepsilon_{it}. \tag{D.1}$$

The prior variance-covariance matrices of A_i and B_i are given by

$$\underline{V}_{A_i} = \text{Var}[\text{vec}(A_i)] = \begin{pmatrix} 0_{k_i^2 \times k_i^2} & 0_{k_i^2 \times k_i k_i^*} \\ 0_{k_i k_i^* \times k_i^2} & \underline{V}_{\Lambda_{i0}} \end{pmatrix}, \tag{D.2}$$

$$\underline{V}_{B_i} = \text{Var}[\text{vec}(B_i)] = \begin{pmatrix} \underline{V}_{\Phi_{i1}} & 0 \\ 0 & \underline{V}_{\Lambda_{i1}} \end{pmatrix}, \tag{D.3}$$

where we assume without loss of generality that covariances between the blocks of coefficients equal zero. Consequently, the expressions for the variance-covariances matrices of $\text{vec}(A_i W_i)$ and $\text{vec}(B_i W_i)$ boil down to

$$\underline{V}_{A_i W_i} = \text{Var}[\text{vec}(A_i W_i)] = (W_i' \otimes I_{k_i}) \underline{V}_{A_i} (W_i' \otimes I_{k_i})', \tag{D.4}$$

$$\underline{V}_{B_i W_i} = \text{Var}[\text{vec}(B_i W_i)] = (W_i' \otimes I_{k_i}) \underline{V}_{B_i} (W_i' \otimes I_{k_i})'. \tag{D.5}$$

Here, for Λ_{i1}, it can easily be seen that the prior variance in country i corresponding to the coefficient associated with country jth endogenous variables is driven by the prior variance-covariance $\underline{V}_{\Lambda_{i1}}$ and the trade weights in W_i.

Equations (D.4) and (D.5) imply that the variance-covariance matrices of G and H in equation (B.2) are given by

$$\underline{V}_G = \begin{pmatrix} \underline{V}_{A_1 W_1} & 0 & \cdots & 0 \\ 0 & \underline{V}_{A_2 W_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \underline{V}_{A_N W_N} \end{pmatrix}, \tag{D.6}$$

$$\underline{V}_H = \begin{pmatrix} \underline{V}_{B_1 W_1} & 0 & \cdots & 0 \\ 0 & \underline{V}_{B_2 W_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \underline{V}_{B_N W_N} \end{pmatrix}, \tag{D.7}$$

which are $k^2 \times k^2$ matrices, respectively. Furthermore, equations (D.6) and (D.7) imply that the covariances between countries equal zero.

Finally, note that when defining conjugate priors locally the corresponding variances only need to be proportional to each other within a given country, since country models are estimated separately in the GVAR framework. This is in stark contrast to a global conjugate prior specification which would require variances of all equations in the system to be proportional to each other.

Posterior simulation and computational issues

Constructing the prior at the local level (and thus leaving the fundamental GVAR structure untouched) facilitates the use of parallel computing. While the majority of the priors described directly come along with analytical posterior solutions, the SSVS prior for example does not. This implies that posterior simulation has to be carried out $N + 1$ times, which might be computationally infeasible. Setting priors locally allows us to fully exploit parallel computing and thus to carry out estimation even in the case posterior distributions are not available in closed form.

In an optimal parallel computing environment the computational speed can be increased c times, when spread across c central processing units (CPUs). However, in reality, only a fraction ϱ of some problem at hand can be parallelized. Amdahl's law (Rodgers, 1985) states that the maximum speedup gained by parallelization is given by $1/\left[\frac{\varrho}{c} + (1 - \varrho)\right]$. In the GVAR case, ϱ is approximately one which implies that the GVAR framework is perfectly suited for exploiting gains from parallelization. This is in contrast to panel VARs (Korobilis, 2015) or large Bayesian VARs (Bańbura *et al.*, 2010) for which computational costs increase more strongly with the dimension of the estimation problem. Thus if the number of available CPUs equals the number of countries, the time needed to estimate a GVAR model using Bayesian methods approximately reduces to the time needed to estimate a single model. In practice, however, overhead costs typically arise. These costs are related to the transportation of the data from the host processor to the nodes. This is negligible relative to the overall estimation time, especially when we have to use simulation based methods. For our present application estimation of a model with $p = p^* = 5$ lags takes between 30 minutes (conjugate specifications) to around two hours (non-conjugate/SSVS prior specifications) on a workstation with eight CPU cores and for 10,000 posterior draws.

Taking into account the computational advantages of such a modelling strategy, posterior inference is done locally, producing draws from the individual country posteriors for all countries in the sample. These draws are transformed using the usual GVAR algebra to produce valid draws from the (joint) global posterior of F and Σ_e, denoted by $p(F, \Sigma_e|\mathcal{D}_T)$, where $\mathcal{D}_T$ denotes the available information set for all countries. Functions of the parameters like forecasts or impulse response functions can be easily calculated using Monte Carlo integration.

Appendix E Forecast Measures

The f-step ahead predictive density of the GVAR model is given by

$$p(x_{t_0+f}|\mathcal{D}_{t_0}) = \int_{\tilde{F}} \int_{\Sigma_e} p(x_{t_0+f}|\tilde{F}, \Sigma_e, \mathcal{D}_{t_0}) p(\tilde{F}, \Sigma_e|\mathcal{D}_{t_0}) d\tilde{F} d\Sigma_e \tag{E.1}$$

where $\mathcal{D}_{T_0}$ denotes the data up to time T_0 and $\tilde{F} = (F_1, \ldots, F_q)$. Note that for conjugate priors, the one-step ahead predictive density is available in closed form. However, for $f > 1$ we either have to resort to numerical methods or turn the problem at hand into a sequence of one-step ahead forecasts. The latter approach, known as the direct method, is employed for all conjugate prior distributions, whereas for non-conjugate distributions we use Monte Carlo integration to approximate the predictive density. This boils down to drawing from the country-specific posteriors and using the algebra outlined in Appendix B to obtain posterior draws of F and Σ_e. We construct the corresponding forecasts by iterating equation B.2 forward and sampling the corresponding errors from $\mathcal{N}(0, \Sigma_e)$. This procedure is repeated n_{sim} times.

As a point estimate, we use the mean of the predictive density described above. Evaluation of the point forecasts is based on the root mean square error (RMSE). The RMSE associated with variable q is given by

$$\text{RMSE}_q = \sqrt{\frac{\sum_{t_0=T_0}^{T-f} (x(q)_{t_0+f}^O - \overline{x}(q)_{t_0+f})^2}{T - T_0 - f + 1}} \tag{E.2}$$

where $x(q)_{t_0}^O$ is the observed data corresponding to the elements in x_t and to variable q. The mean of the f-step ahead predictive density of variable q is denoted by $\overline{x}(q)_{t_0+f}$.

The log predictive score (LPS), as given in e.g. Geweke & Amisano (2010), is the predictive density given by equation (E.1) evaluated at the realized outcome.

$$\text{LPS}(x_{t_0+f}^O|\mathcal{D}_{t_0}) = \log p(x_{t_0+f} = x_{t_0+f}^O|\mathcal{D}_{t_0}) \tag{E.3}$$

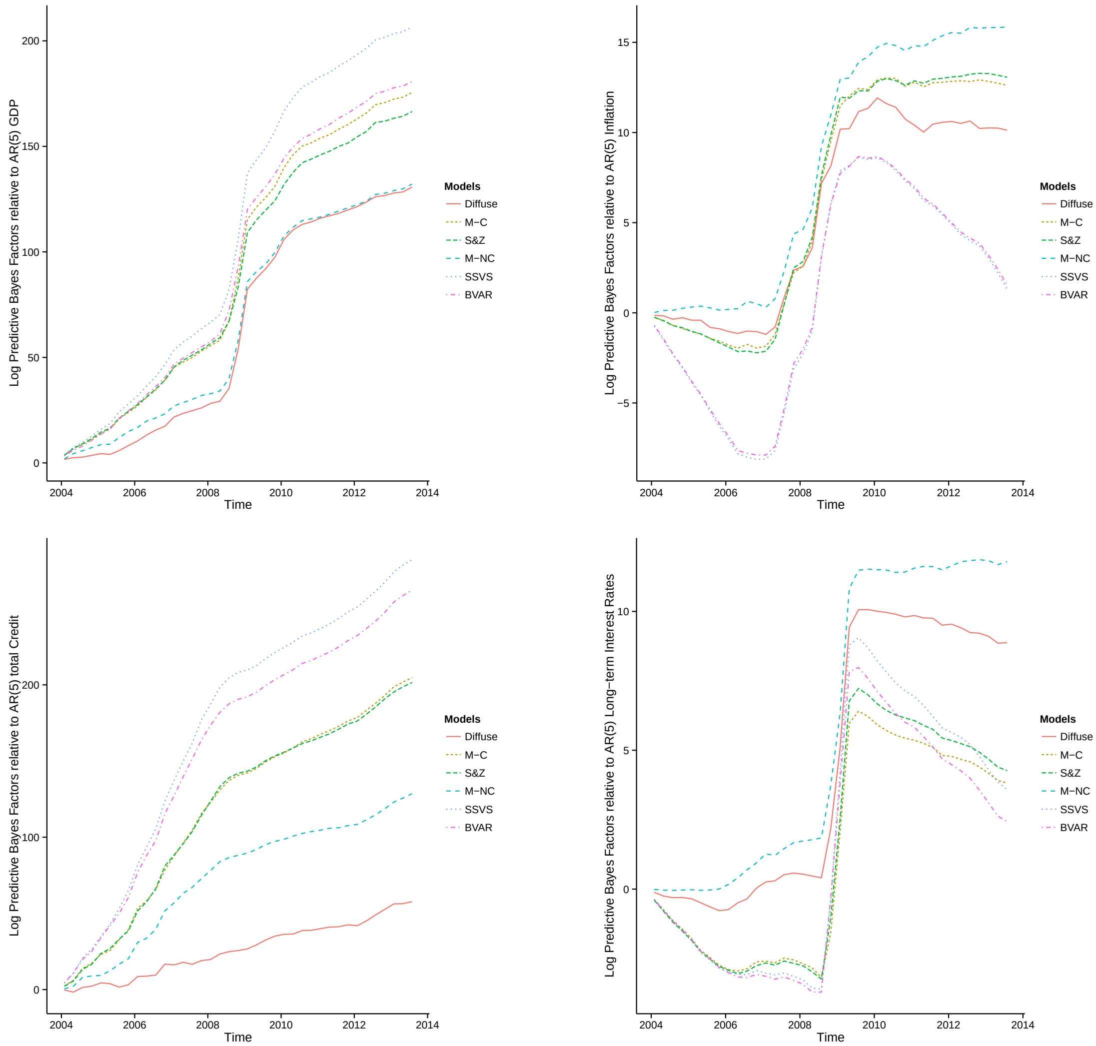
As noted above, for $f > 1$ equation (E.3) has no closed form solution. Following Adolfson *et al.* (2007) we approximate the LPS using a multivariate normal density which is evaluated with posterior mean estimates from the predictive density. This second-order approximation is given by

$$\begin{aligned}\widehat{LPS}(x_{t_0+f}^O|\mathcal{D}_{t_0}) \approx & -0.5[k \log(2\pi) + \log|\overline{\Omega}_{t_0+f|t_0}| \\ & + (x_{t_0+f}^O - \overline{x}_{t_0+f|t_0})' \overline{\Omega}_{t_0+f|t_0}^{-1} (x_{t_0+f}^O - \overline{x}_{t_0+f|t_0})],\end{aligned} \tag{E.4}$$

where $\overline{\Omega}_{t_0+f|t_0}$ denotes the mean of the f-step ahead predictive variance-covariance matrix. We present variable-specific log predictive scores calculated by integrating out the effect of other variables in the system. Under the assumption of multivariate normality, the LPS associated to variable q can be calculated by deleting the rows (and columns) corresponding to variable $g \neq q$ from $x_{t_0+f}^O, \overline{x}_{t_0+f|t_0}$ and $\overline{\Omega}_{t_0+f|t_0}$, respectively.

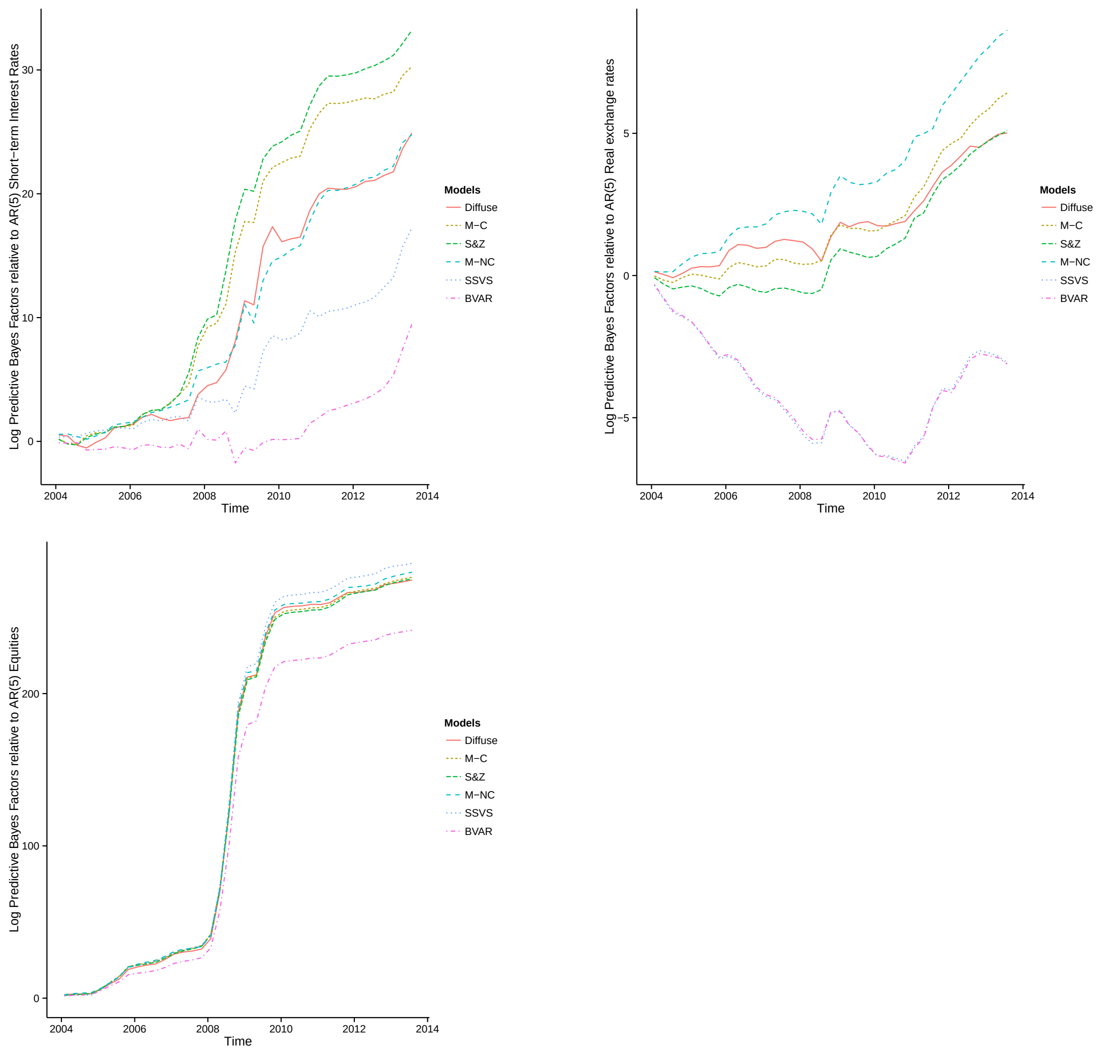
Appendix F Forecast Ability: Changes Over Prediction Sample

Figure F.1: LPS relative to AR(5) benchmark over forecasting period: GDP, inflation, total credit and long-term interest rates



Notes: Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed.

Figure F.2: LPS relative to AR(5) benchmark over forecasting period: Short-term interest rates, real exchange rates and equities



Notes: Diffuse stands for the model estimated using maximum likelihood, M-C denotes the GVAR with the conjugate variant of the Minnesota prior, S&Z refers to a GVAR estimated using a weighted average of the conjugate priors as in Sims & Zha (1998), M-NC stands for the variant of the Minnesota prior with a non-random variance-covariance matrix of the error term, SSVS denotes the GVAR estimated using the SSVS prior, and BVAR denotes forecasts based on separate country VARs with the SSVS prior employed.