Density Forecasting using Bayesian Global Vector Autoregressions with Common Stochastic Volatility

Florian Huber

Oesterreichische Nationalbank (OeNB)

November 11, 2014

Abstract

This paper puts forward a Bayesian Global Vector Autoregressive Model with Common Stochastic Volatility (B-GVAR-CSV). We assume that country specific volatility is driven by a single latent stochastic process, which simplifies the analysis and implies significant computational gains. Apart from computational advantages, this is also justified on the ground that the volatility of most macroeconomic quantities considered in our application tends to follow a similar pattern. Furthermore, Minnesota priors are used to introduce shrinkage to cure the curse of dimensionality. Finally, this model is then used to produce predictive densities for a set of macroeconomic aggregates. The dataset employed consists of quarterly data spanning from 1979:Q1 to 2013:Q4 and includes 36 economies. Our results indicate that allowing for stochastic volatility influences the accuracy along two dimensions: First, it helps to increase the overall predictive fit of our model. This result can be seen for all variables under scrutiny, most notably for real GDP, inflation, exchange rates and equity prices. Second, it helps to make the model more resilient with respect to outliers and economic crises. This implies that when evaluated over time, the log predictive scores tend to show significantly less variation as compared to homoscedastic models.

JEL Classification: C32, F44, E32, E47.
Keywords: Density Forecasting, Stochastic Volatility, Global vector autoregressions.

1 Introduction

Recent episodes of rising volatility of several key macroeconomic quantities revealed that most models employed in policy institutions failed to deliver reliable forecasts under such circumstances. This stems from the fact that practitioners remained largely confined to simple linear models which do not account for structural changes in the behavior of the underlying variables. Two reasons are worth mentioning why the majority of applied researcher still stick to linear models. First, estimation is easy and numerical optimization is often unnecessary. As a consequence, they are easy to implement using standard statistical software packages. Second, linear models are easy to interpret and understand, which makes them valuable for the majority of practitioners. However, the recent global turmoil has proved that more flexible models are needed to fully capture the complex dynamics arising in macroeconomics and finance. Especially for highly volatile financial time series non-linear models are needed to fully capture sudden shifts in volatility commonly observed in financial markets.

Several studies provided evidence for a sudden increase of volatility in industrialized economies after experiencing decades of relatively stable and low volatility of macroeconomic fundamentals. Linear models, like vector autoregressive models (VARs), which have been performing quite well up to the mid 2000s suddenly failed to produce reliable predictions. Ignoring the dynamic behaviour of volatility led to predictive densities which are either too narrow or too wide, resulting in inflated confidence bounds and poorly estimated probabilities for tail events. Thus it might be necessary to account for heteroscedasticy by means of more flexible specifications of the variance covariance matrix. A plethora of studies emphasized the usefulness of such stochastic volatility specifications in terms of point- and density forecasts. Giordani and Villani (2010), Clark (2011) and Carriero et al. (2012) all highlight the substantial increases in forecasting accuracy by using SV specifications. Such gains in accuracy typically directly translate into better prediction intervals produced in central banks and other policy institutions, underlining the practical relevance of this approach.

Despite the fact that stochastic volatility VARs introduce additional flexiblity when it comes to macroeconomic modeling, computational needs also increase substantially. Additionally, VARs typically suffer from the well-known "curse of dimensionality" which implies that the model overfits the data. This could translate into weak out-of-sample prediction performance. Thus, Bayesian methods are needed to obtain reliable estimates and impose shrinkage on the parameters. Furthermore, allowing for flexible stochastic volatility specifications in VARs typically leads to non-conjugate situations where forward-filtering-backward-sampling methods (FFBS) are required. This bounds the analysis usually to small- to medium scale models. Especially in forecasting applications, it is of interest to allow for high dimensional models to exploit information originating from other variables or other countries. Several recent contributions aimed for making the estimation of large scale models feasible and still preserve the flexibility of non-linear models. Koop and Korobilis (2013) draw on ideas from the dynamic model averaging literature and utilize forgetting factors to reduce the computational burden. In another contribution, Carriero et al. (2012) allow for a simplified version of stochastic volatility, where it is assumed that the volatility of the whole system is driven by a single, latent process. This assumption preserves the conjugacy of the prior and permits a convenient Kronecker structure of the likelihood. This implies significant computational gains using such a simplified prior structure.

In terms of achieving shrinkage in large scale macroeconometric models, the global vector autoregressive model (GVAR) put forward by Pesaran et al. (2004) proved to be a convenient way of reducing the dimensionality of the estimation problem. Several contributions outlined the usefulness of such large scale models to perform forecasting (Pesaran et al., 2009; Greenwood-Nimmo et al., 2012; Crespo Cuaresma et al., 2014) or impulse response analysis (Pesaran et al., 2007; Dees et al., 2007). One disadvantage is, however, that frequentist estimation of GVAR models does not cure the curse of dimensionality at the local level. This implies that even though the dimensionality of the problem is reduced considerably, local models might also suffer from severe overfitting. Crespo Cuaresma et al. (2014) proposed a Bayesian variant of the GVAR and evaluated its predictive performance in a forecasting horse race. It is shown that Bayesian shrinkage, in addition to the restrictions imposed by the GVAR, helps to improve point and density forecasts for all variables under consideration.

In the present paper we propose a Bayesian variant of the GVAR which allows for a time varying variance-covariance structure (B-GVAR-CSV) in the spirit of Carriero et al. (2012). This implies that the local models, which are stacked in a second stage to yield the global model, are driven by a single latent stochastic process which governs the country specific log-volatities. That means in each country model, that consists of several single equations for the macrovariables at hand, we model one stochastic volatility process as opposed to having stochastic volatility modelled in each equations separately. As a consequence, the global system, which comprises of the $N + 1$ local systems, is driven by $N + 1$ local latent factors. The contributions of this paper are threefold. First, the possibility to allow for stochastic volatility in the GVAR is introduced. A first attempt to model time varying volatilities has been recently adopted in Cesa-Bianchi et al. (2014), where a satellite model for the volatility process is introduced. However, in this paper we take a more coherent approach and model stochastic volatility for each country separately. Second, we propose a simple and efficient algorithm to estimate the local models. In particular, sampling the log volatilities is done using the algorithm outlined in Kastner and Frühwirth-Schnatter (2013). As compared to the estimation of standard Bayesian VARs, this method is extremely fast, resulting only in marginally higher computational needs. Finally, we use the B-GVAR-CSV to forecast several key macroeconomic quantities and evaluate their predictive densities. Our results suggest that the introduction of stochastic volatility leads to more precise density forecasts as measured by log predictive scores for all variables under scrutiny at all time horizons considered, where especially for GDP, inflation, exchange rates and equities the GVAR with CSV consistently outperforms its peers.

This paper is structured as follows. Section 2 introduces the econometric framework employed while Section 3 discusses prior setups and the Markov-Chain Monte Carlo (MCMC) algorithm. Section 4 presents the dataset and the results of the density forecasting exercise. Finally, the last Section concludes.

2 The B-GVAR with Stochastic Volatility

2.1 From local to global: The GVAR Model

The main building block of the GVAR model put forward by Pesaran et al. (2004) are the local macroeconomic models. More specifically, we assume that domestic variables are modeled using a standard VAR with exogenous regressors (VARX*). A typical VARX* model for country $i = 0, ..., N$ is then given by

$$x_{i,t} = a_{i0} + a_{i1}t + \sum_{s=1}^{p} A_{is}x_{i,t-s} + \sum_{r=0}^{q} B_{ir}x^*_{i,t-r} + \varepsilon_{i,t} \tag{2.1}$$

where $x_{i,t}$ denotes a $k_i \times 1$ vector of endogenous variables measured in country i at time t and a_{ij} $(j = 1, 2)$ are k_i-dimensional vectors corresponding to the constant and the trend. Furthermore, A_{is} denotes the $k_i \times k_i$ coefficient matrix corresponding to the s'th lag of the endogenous variables. This part of equation (2.1) captures domestic dynamics. The $k^*_i \times 1$ vector $x^*_{i,t}$ denotes the so-called *weakly exogenous variables*, which are defined as

$$x^*_{i,t} = \sum_{j \neq i}^{N} \omega_{i,j} x_{j,t} \tag{2.2}$$

where $\omega_{i,j}$ denotes the weight between countries i and j and $\sum_{j \neq i}^{N} \omega_{i,j} = 1$. The $k_i \times k_i^*$ coefficient matrix related to $x^*_{i,t-r}$ is given by B_{ir}. Note that the discrimination between strictly exogenous and weakly exogenous is crucial because the latter will become effectively endogenous once the model is solved. Finally, $\varepsilon_{i,t} \sim \mathcal{N}(0, \Sigma_{i,t})$ is the usual vector white noise process. The dynamics of the variance covariance matrix, following Carriero et al. (2012), are assumed to be driven by a single latent stochastic process $h_{i,t}$. More specifically, we assume that $\Sigma_{i,t}$ evolves according to

$$\Sigma_{i,t} = \exp\left(h_{i,t}/2\right) \times \Sigma_i \tag{2.3}$$

$$h_{i,t} = \eta_i + \xi_i (h_{i,t-1} - \eta_i) + \sigma_i \epsilon_{i,t} \tag{2.4}$$

$$\epsilon_{i,t} \sim \mathcal{N}(0, 1) \tag{2.5}$$

where we assume that $\xi_i \in (-1, 1)$. This implies that the stochastic process which governs the log-volatility is mean reverting. It would be possible to assume that the log-volatility follows a random walk process. However, this implies that the log-volatility is unbounded in the limit (for a discussion on whether to model log-volatities as stationary or non-stationary see Eisenstat and Strachan, 2014). Such behavior is ruled out using this more general specification. Moreover, to achieve identification we simply set $h_{i,0} = 0$. This completes the discussion of the local models.

The set of $N + 1$ country models can be connected using a suitable $(k_i + k_i^*) \times k$ (with $k = \sum_{i=0}^{N} k_i$) country-specific linking matrices W_i. Following Pesaran et al. (2004) it is easy to show that the country-specific models can be stacked to yield the global representation of the model

$$G x_t = a_0 + a_1 t + \sum_{s=1}^{p^*} \Gamma_s x_{t-s} + u_t \tag{2.6}$$

where G and Γ are complex functions of the parameters of the individual country models, $p^* = \max(p, q)$ and $u_t \sim \mathcal{N}(0, \Sigma_t)$ with

$$\Sigma_t = \begin{pmatrix} \exp(h_{0,t}/2) \times \Sigma_0 & 0 & \cdots & 0 \\ 0 & \exp(h_{1,t}/2) \times \Sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \exp(h_{N,t}/2) \times \Sigma_N \end{pmatrix} \tag{2.7}$$

which implies that the log-volatility of the global system is governed by $N + 1$ latent stochastic processes. Furthermore, note that this assumption implies that until now, the cross-country covariances are set equal to zero. Solving the model in (2.6) for x_t yields

$$x_t = b_0 + b_1 t + \sum_{s=1}^{p^*} \Lambda_s x_{t-r} + e_t \tag{2.8}$$

with $b_0 = G^{-1}a_0, b_1 = G^{-1}a_1, \Lambda_s = G^{-1}\Gamma_s$ and $e_t = G^{-1}u_t$. This implies that $e_t \sim \mathcal{N}(0, \Omega_t)$ with $\Omega_t = G^{-1}\Sigma_t G^{-1'}$, which is in general not block-diagonal.

The GVAR model in equation (2.8) resembles a standard, high-dimensional VAR. Thus we can use (2.8) to produce forecasts, impulse responses or forecast error variance decompositions.

2.2 General Prior Setup

To conduct Bayesian inference we have to specify prior distributions for all parameters in the model. Following Crespo Cuaresma et al. (2014), this is done at the individual country level. For further convenience let us rewrite the local models as

$$x_i = A_i z_i + B_i z_i^* + \varepsilon_i \tag{2.9}$$

where x_i is a $T \times k_i$ matrix of stacked data and $A_i = (a_{i0}, a_{i1}, A_{i1}, \ldots, A_{ip})$ is a $k_i \times (2 + k_i p)$ matrix of stacked coefficients. Additionally, z_i is a $T \times k_i p$ matrix of explanatory variables where the jth row is $z_{ij} = (1', t', x'_{i,j-1}, \ldots, x'_{i,j-p})'$. Likewise, z_i^* is a $T \times (k_i^* q + 2)$ matrix with typical row j given by $z_{ij}^* = (x^{*'}_{i,j-1}, \ldots, x^{*'}_{i,j-p})'$.

The prior setup for the coefficients in country i is then given by

$$\begin{aligned}
\text{vec}(A_i)|\Sigma_i &\sim \mathcal{N}(\text{vec}(\underline{A}_i), \Sigma_i \otimes \underline{V}_{A_i}), \\
\text{vec}(B_i)|\Sigma_i &\sim \mathcal{N}(\text{vec}(\underline{B}_i), \Sigma_i \otimes \underline{V}_{B_i}), \\
\Sigma_i^{-1} &\sim \mathcal{W}(\underline{v}_i, \underline{S}_i^{-1}) \\
\eta_i &\sim \mathcal{N}(\underline{\eta}_i, \underline{V}_\eta) \\
\frac{\xi_i + 1}{2} &\sim \mathcal{B}(a_0, b_0) \\
\sigma_i &\sim \mathcal{G}(1/2, 1/2B_\sigma) \\
\theta_i &\sim \mathcal{G}(c_0, c_1)
\end{aligned} \tag{2.10}$$

Note that we assume prior dependence between A_i, B_i and Σ_i, which implies that we can exploit a Kronecker structure for the likelihood. The Kronecker structure, as mentioned in Carriero et al. (2012), leads to large increases in computational efficiency, especially when the number of endogenous variables is increased at the local level.

Several choices for $\underline{A}_i, \underline{B}_i$ and the corresponding variance-covariance matrices $\underline{V}_{A_i}$ and $\underline{V}_{B_i}$ are possible. However, we will restrict our analysis to the well-known Minnesota prior, which shrinks the system towards a naïve random walk process. Exact details for the implementation can be found below. The prior on the time-invariant part of the precision Σ_i^{-1} is of standard Wishart form with prior degrees of freedom $\underline{v}_i$ and scale matrix $\underline{S}_i^{-1}$. Due to the conjugacy of the prior setup discussed above it is possible to implement it solely by means of so-called dummy observations. This implies that artificial data observations are concatenated to the actual data and exploit Theil-Goldberger mixed estimation (Theil and Goldberger, 1961) to obtain the posterior moments for A_i, B_i and Σ_i.

Furthermore, for the level η in the log-volatility equation we impose a normal prior with mean $\underline{\eta}_i$ and variance $\underline{V}_{\eta_i}$. Following Kastner and Frühwirth-Schnatter (2013) we impose a beta prior on the persistence parameter ξ_i. Formally, the prior density is

$$p(\xi_i) = \frac{1}{2B(a_0, b_0)} \frac{(1+\xi_i)}{2}^{(a_0-1)} \frac{(1-\xi_i)}{2}^{(b_0-1)} \tag{2.11}$$

where $B(a_0, b_0)$ denotes the beta function. The support of this distribution is the unit ball, which implies stationarity of the log-volatility process. The prior mean and variance are equal to

$$E(\xi_i) = \frac{2a_0}{a_0 + b_0} - 1$$
$$\text{Var}(\xi_i) = \frac{4a_0 b_0}{(a_0 + b_0)^2(a_0 + b_0 + 1)}$$

Note that if $\frac{2a_0}{a_0+b_0} < 1$, the prior mean is negative. Obviously, this case would coincide with setting $b_0 > a_0$. A positive prior mean would correspond to the case when $a_0 > b_0$. For typical datasets arising in macroeconomics the exact choice of the hyperparameters a_0 and b_0 is quite influental, due to the short time series available. We impose a non-conjugate gamma prior for σ_i. This choice has the advantage that it does not bound σ_i away from zero and increases sampling efficiency considerably. Further details can be found in Frühwirth-Schnatter and Wagner (2010) and Kastner and Frühwirth-Schnatter (2013).

Finally the prior section is concluded by imposing a prior on the hyperparameters in the spirit of Giannone et al. (2012). We store the set of hyperparameters controlling the tightness of the prior on A_i, B_i and Σ_i in a vector θ_i and impose a Gamma prior. In the present setting it is crucial to allow for multiple hyperparameters controlling the tightness of the prior on the endogenous and weakly exogenous part of the model, separately.

We will discuss the exact choice of the hyperparameters at length in section 3.

2.3 Posterior Distributions

For the present application the conditional posteriors for A_i, B_i and Σ_i are of a well-known form. Namely multivariate normal distributions for A_i and B_i and a inverse-Wishart distribution for Σ_i. This implies that those parts can be sampled using a Gibbs sampling scheme. Drawing the parameters of the stochastic volatility equation is then done following Kastner and Frühwirth-Schnatter (2013) using a ancillarity-sufficiency interweaving strategy (ASIS).

Let us define some additional notation used to describe the posterior moments of A_i, B_i and Σ_i. Assume that the data for each country i is stored in

$$Z_{i,t} = (1, t, x_{i,t-1}, ..., x_{i,t-p}, x^*_{i,t}, ..., x^*_{i,t-q}). \tag{2.12}$$

Defining a $k_i \times K_i$ (where $K_i = 2 + k_i p + k_i^* q$) matrix of stacked coefficients $\Psi_i = (A_i, B_i)$ allows us to further simplify the model in (2.9) as

$$x_{i,t} = \Psi_i Z_{i,t}' + \varepsilon_{i,t}. \tag{2.13}$$

Stochastic volatility is introduced by dividing $Z_{i,t}$ and $x_{i,t}$ by $\exp(h_{i,t}/2)$:

$$\tilde{x}_{i,t} = \exp(-h_{i,t}/2)\ x_{i,t}$$
$$\tilde{Z}_{i,t} = \exp(-h_{i,t}/2)\ Z_{i,t}.$$

In what follows, we denote the corresponding full-data matrices of $\tilde{x}_{i,t}$ and $\tilde{Z}_{i,t}$ as $\tilde{x}_i$ and $\tilde{Z}_i$.

Given $\tilde{x}_{i,t}$ and $\tilde{Z}_{i,t}$ it is straightforward to describe the conditional posterior distributions for Ψ_i and Σ_i:

$$\text{vec}(\Psi_i)|\Sigma_i, h_{i,t}, \eta_i, \xi_i, \mathcal{D}_T \sim \mathcal{N}(\text{vec}(\overline{\Psi}_i), \Sigma_i \otimes \overline{V}_{\Psi_i}) \tag{2.14}$$
$$\Sigma_i^{-1}|\Psi_i, \eta_i, \xi_i, \mathcal{D}_T \sim \mathcal{W}(\overline{v}_i, \overline{S}_i) \tag{2.15}$$

where $\mathcal{D}$ is denotes all available data up to time T. As mentioned above, the natural conjugate prior can be implemented through dummy observations. Stacking a set of dummy observations $\underline{x}_i, \underline{Z}_i$ with the actual data yields $\overline{x}_i = (\tilde{x}_i', \underline{x}_i')'$ and $\overline{Z}_i = (\tilde{Z}_i', \underline{Z}_i')'$.

The posterior quantities for Ψ_i (and thus the corresponding moments for A_i and B_i) are readily obtained using mixed estimation. Formally, the posterior mean and variance for Ψ_i are given by

$$\overline{\Psi}_i = (\overline{Z}_i'\overline{Z}_i)^{-1}\overline{Z}_i'\overline{x}_i \tag{2.16}$$
$$\overline{V}_{\Psi_i} = (\overline{Z}_i'\overline{Z}_i)^{-1} \tag{2.17}$$

For the variance-covariance matrix the posterior degrees of freedom and scale matrix are given by

$$\overline{v}_i = \underline{v}_i + T \tag{2.18}$$
$$\overline{S}_i = (\overline{x}_i - \overline{Z}_i\overline{\Psi}_i')'(\overline{x}_i - \overline{Z}_i\overline{\Psi}_i'). \tag{2.19}$$

The components of the log-volatility equations possess posterior distributions of no-well known form, which precludes simple Gibbs sampling schemes. The same holds true for the posterior of θ_i. However, due to the (conditional) conjugacy of our model the marginal likelihood is available in closed form. This permits us to set up a simple Metropolis Hastings step to obtain posterior draws of θ_i.

3 Implementation & Estimation

3.1 Prior Implementation

Up to this point we have remained silent on the exact prior settings. For the B-GVAR-CSV we utilize a standard implementation of the well-known Minnesota prior to achieve shrinkage at the local level. Following Karlsson (2012) this implies setting the prior moments according to

$$[\underline{\Psi}_i]_{lj} = \begin{cases} \underline{\delta}_{lj} & \text{for the first, own lag of an endogenous variable} \\ 0 & \text{in all other cases} \end{cases} \tag{3.1}$$

$$\underline{V}_{\Psi_i,lj} = \begin{cases} \frac{\alpha_{i1}}{(r^2 \varsigma_{ij})^2} \text{ for coefficients on lag } s = 1, ..., p \text{ of variable } j \\ \frac{\alpha_{i2}}{(\varsigma^*_{ij}(1+k))^2} \text{ for coefficients on lag } k = 0, ..., q \text{ of the weakly exogenous variables } j \\ \underline{\alpha}_{i3} \text{ for the deterministic part of the model} \end{cases} \tag{3.2}$$

where $\underline{\delta}_{lj}$ is set to one for all variables except for inflation, where it is set equal to 0.2. Note that the ς_{ij} refer to the standard deviations obtained by running univariate autoregressions in a given country and ς^*_{ij} are obtained by running univariate autoregressions on the weakly exogenous variables in country i. To match the moments of the Minnesota prior we follow Banbura et al. (2010a) and Crespo Cuaresma et al. (2014)

and use the following set of dummy observations

$$
\underline{x}_i = \begin{pmatrix} 0_{2\times k_i} \\ \dotfill \\ \text{diag}(\underline{\delta}_{i1}\varsigma_{i1}, \ldots, \underline{\delta}_{ik_i}\varsigma_{ik_i})/\alpha_{i1} \\ 0_{k_i(p-1)\times k_i} \\ \dotfill \\ 0_{k_i^*(q+1)\times k_i} \\ \dotfill \\ \text{diag}(\varsigma_{i1}, \ldots, \varsigma_{ik_i}) \end{pmatrix}
$$

$$
\underline{Z}_i = \begin{pmatrix} I_2\frac{1}{\alpha_3} & & 0_{2\times(K_i-2)} \\ \dotfill & & \\ 0_{k_ip\times 2} & J_p \otimes \text{diag}(\sigma_{i1}, \ldots, \varsigma_{ik_i})/\alpha_1 & 0_{k_ip\times(K_i-k_ip-2)} \\ \dotfill & & \\ 0_{(k_i^*(q+1))\times(k_ip+2)} & & J_q \otimes \text{diag}(\varsigma_{i1}^*, \ldots, \varsigma_{ik_i^*}^*)/\alpha_{i2} \\ \dotfill & & \\ & 0_{k_i\times K_i} & \end{pmatrix}, \tag{3.3}
$$

where $0_{n,q}$ denotes a $n \times q$ dimensional zero matrix and $J_p = \text{diag}(1, 2, ..., p)$. Finally, $J_q = \text{diag}(1, 2, ..., q + 1)$ is constructed in the same way. The first two blocks of equation (3.3) implement the prior on the domestic part of the model in (2.1). The third block is related to the prior on the weakly exogenous variables while the final block implements the prior on Σ_i.

The hyperparameters $\theta_i = (\underline{\alpha}_{i1}, \underline{\alpha}_{i2}, \underline{\alpha}_{i3})$ are sampled using a simple Metropolis Hastings step outlined above. This prior setup has several implications. First, the endogenous part of the model is shrunk towards a stylized prior model, usually assumed to be a (multivariate) random walk. This is justified by the fact that the random walk has a proven track record in forecasting applications. Second, the hierarchical nature of our prior setup allows us to find an appropriate degree of shrinkage for the weakly exogenous block of the model. Especially in our setup where we are interested in judging the relative importance of the weakly exogenous variables this proves to be a great advantage as compared to settings which are not hierarchical. Moreover, allowing for different hyperparameters across countries also allows us to incorporate cross-country differences in a flexible fashion. Variants of this prior setup has been used extensively in the literature with great success in forecasting applications (see, for example, Kadiyala and Karlsson, 1997; Bańbura et al., 2010b; Koop, 2013, among others).

The hyperparameters for the log-volatility equation are set as follows. For the level η_i we set the mean $\underline{\mu}_\eta = 0$ and the variance $\underline{V}_\eta = 10^2$. This implies a non-informative prior distribution on η_i. For the persistence parameter ξ_i we set $a_0 = 5$ and $b_0 = 1.5$, resulting in a prior mean of around 0.54 and standard deviation 0.31. Finally, for σ_i we set the $B_\sigma = 1$. For our application, the exact choice of B_σ is not critical, as long as it is set sufficiently large.

3.2 Posterior Simulation: The MCMC algorithm

The MCMC algorithm for country i works as follows.

1. Draw $\Psi_i | \Sigma_i, h_{i,t}, \eta_i, \xi_i, D$ from $\mathcal{N}(\overline{\Psi}_i, \overline{V}_{\Psi_i})$
2. Draw $\Sigma_i^{-1} | \Psi_i, \eta_i, \xi_i, D$ from $\mathcal{W}(\overline{v}_i, \overline{S}_i)$
3. Draw the parameters of the log-volatility equations and $h_i = (h_{i,0}, ..., h_{i,T})'$ using the AWOL-sampler described in Kastner and Frühwirth-Schnatter (2013)

Steps (1) - (2) can be implemented in a straightforward fashion by sampling from Normal and Wishart distributions, respectively. The parameters of the stochastic volatility equation are updated using the ASIS algorithm. The main idea behind this interweaving strategy is that depending on which parametrization we use for the stochastic volatility process (i.e. whether we use the centered parameterization shown above or a non-centered parameterization), sampling efficiency is increased by combining "the best of two worlds".

Sampling $h_{i,t}$ and the parameters of the log-volatility equation is done in three steps. In the first step, we draw the latent volatilities all without a loop. Conditional on the other parameters, h_i is multivariate normal, where the first-order autoregressive nature makes it possible to write this density in terms of a tridiagonal-precision matrix Ω_h, which makes it possible to sample h_i all without a loop. Note that this is true for the centered and non-centered parameterizations.

In the second step, we sample the parameters of the stochastic volatility equation using Metropolis Hastings steps. Centered and non-centered parameterizations require a 2- and 3-block sampler respectively.

Finally, note stochastic volatility models are usually non-Gaussian. This implies that the innovations are $\log \chi(1)^2$ distributed. Standard practice in the literature is to circumvent this problem by using a mixture of Gaussian distributions to approximate the $\log \chi(1)^2$-distribution. In this case, we have to sample indicators which govern the normal distribution to use through inverse transform sampling. For further details we refer the reader to Kastner and Frühwirth-Schnatter (2013). These steps are implemented using the `stochvol` package, which is available on CRAN.

Treating the local models as separated estimation problems facilitates the usage of parallel computing. That is, we can distribute the estimation of individual country model across different CPU cores. This implies that in our case, if we would have $N + 1$ available CPU cores on a cluster it would take approximately the same time as the estimation of a single VARX* model.[1]

This delivers draws from the country specific joint posterior. However, interest centers on the posterior for the global model, which can be readily obtained using draws from $p(\Psi_i|\bullet)$, $p(\Sigma_i|\bullet)$ and $p(h_i|\bullet)$ for all $i = 0, ..., N$, where the dot indicates conditioning on all other coefficients and the data. Using the algebra described in Pesaran et al. (2004) allows us to transform draws from the country-specific posteriors to obtain valid draws from the global posterior of $\Upsilon = (b_0, b_1, \Lambda_1, \dots, \Lambda_{p^*})$ and Ω_t.

4 Empirical Forecasting Application

4.1 Data Overview and Forecasting design

We use the same dataset as Crespo Cuaresma et al. (2014), which consists of 36 countries covering a broad set of developed and developing economies. Further information on the countries included can be found in the appendix. The dataset spans the time period from 1979Q1 to 2013Q4, which are 137 quarterly observations.

Following the literature on GVARs (Pesaran et al., 2004; Dees et al., 2007; Pesaran et al., 2009), we use a fairly standard set of macro aggregates in our individual country models. Table A.1 provides a brief description of the variables employed. Note that most variables are available for nearly all countries in our dataset, with the exception of long-term bond yields. As a global control variable we include the price of oil. For a more complete description of the data see Crespo Cuaresma et al. (2014).

The weakly exogenous part in the local model is symmetric with the exception of the real exchange rate vis-a-vis the US dollar. The weakly exogenous exchange rate is only included in the US country model. Moreover, we include three lagged endogenous and weakly-exogenous variables in each country model. In light of the length of our data set, this particular choice seems to be appropriate.

For forecasting purposes we have to simulate the predictive density of the global model. The h-step ahead predictive density at time t is given by

$$p(x_{t+h}|x_{1:t}) = \int_{\Upsilon} \int_{\Omega_{t+h}} p(x_{t+h}|\Upsilon, \Omega_{t+h}, x_{1:\tau}) p(\Upsilon, \Omega_{t+h}|x) \; d\Upsilon d\Omega_{t+h}, \tag{4.1}$$

where $x_{1:t}$ denotes the data contained in x up to time t.

Due to the fact that the posterior of Υ and Ω_t is available we can obtain (4.1) using Monte Carlo integration. To evaluate the predictive capabilities in terms of density forecasts

[1] One would also have to take into account overhead times from distributing the data to the nodes, but compared to the whole estimation time this part is negligible.

we utilize the well known log predictive score (LPS), given by

$$\text{LPS} = \sum_{t=t_0}^{T-h} \log p(x_{t+h} = x^o_{t+h} | x_{1:t}) \tag{4.2}$$

where x^o_{t+h} denotes the observed outcome of x at time $t+h$. The LPS provides guidance on the overall predictive fit of the model. As an quadratic approximation for the LPS we follow Adolfson et al. (2007) and use a Gaussian approximation to the LPS.

The LPS measures the out-of-sample predictive capabilities of our model when used to predict the whole system jointly. However, we are more interested in differences in terms of predictive accuracy across the different variables. Consequently to investigate the joint density of some variable of interest we simply integrate out all other variables. In light of the quadratic approximation this can be done in a straightforward fashion by exploiting the properties of the multivariate Normal distribution.

Judgement of the models is done exclusively based on log predictive scores. The reasons for this are twofold. First, model evaluation based on point forecasts typically disregards the uncertainty surrounding the predictions. Second, the point of stochastic volatility specification is to increase the robustness of the model with respect to changing magnitudes of economic shocks. Thus, judging the models by their point forecasts would result in a situation where the variability is averaged out, indicating that the value added of a stochastic volatility specification delivers is effectively purged out.[2]

The forecasting design employed is the following: We start with an initial estimation period $t_0 = 1979Q4$ to $t_1 = 2004Q1$ and simulate the h-step ahead predictive density for $p(x_{t_1+h}|x_{1:t})$. After retrieving density estimates we proceed by setting $t_1 = t_{1+h}$ and again producing the h-step ahead predictive density. This procedure is repeated until we reach $t_1 = T - h$.

Our goal is to prove empirically that allowing for a simplified form of stochastic volatility leads to pronounced gains in terms of density forecasting accuracy. Thus as a competitor we include the GVAR coupled with the same (hierarchical) Minnesota prior.

4.2 Results

Table 1 presents the results of our forecasting exercise. The numbers refer to the (average) differences in log predictive scores between the GVAR with a common stochastic volatility specification coupled with a Minnesota prior and a homoscedastic GVAR with the same

[2]We have computed point forecasts and the results corroborate this statement: The differences in the accuracy of the point forecasts between the homoscedastic Minnesota prior GVAR and the CSV variant are quite small.

Minnesota prior. Numbers greater than zero indicate outperformance in terms of the CSV specification vis-á-vis the baseline homoscedastic GVAR.

[Table 1 about here.]

For real GDP (y), the outperformance in terms of log predictive score is large at the one- and four quarter ahead time horizon. Table 1 suggests that especially at short time horizons, allowing for stochastic volatility improves the precision of the predictive densities for output considerably. The differences arise due to a strong performance of the CSV specification in the recent financial crisis. As can be seen in Figure 1a the differences in log predictive scores for output tend to vary strongly over time. One remarkable pattern is that while the CSV specification performs poorly relative to the baseline model prior to the recent financial crisis, it recuperates strongly in the crisis, providing evidence that stochastic volatility helps to robustify the analysis with respect to economic outliers. This pattern can be seen for both time horizons considered, where the evidence is even stronger at the one year ahead time horizon.

The prediction gains for CPI inflation (Δp) again favour the CSV specification by large margins. Inspection of the evolution of the differences in log scores reveals that in the period before the financial crisis, allowing for stochastic volatility helps to improve the precision of the forecasts over the time frame from 2004Q1 to 2006Q3. While the time period between 2007 and the onset of the financial crisis was characterized by a sudden drop in forecasting accuracy of the GVAR-CSV relative to the standard GVAR, the CSV specification quickly recovered and led to steady gains. These gains are especially pronounced in the crisis. Note that after the crisis the log score differences have been stable and positive until the end of 2013. This point towards changes in the underlying macroeconomic conditions, further reinforced by the crisis of 2008/09.

The CSV specification again possesses advantages when used to conduct one quarter ahead forecasts of the short-term interest rates (i_S). For the one year ahead forecasts the log predictive scores between Minnesota and CSV are comparable, with minor advantages for the latter variant of the model. However note that the prediction gains are somewhat less pronounced for both horizons. Visual inspection of the corresponding differences in log scores reveals that after the crisis hit the global economy, the CSV specification improved upon the benchmark by a large extent. For the time periods before and after the crisis the prediction gains are mostly rather small or even negative for some periods (especially between 2006 and the second half of 2007).

For long-term interest rates (i_L) the CSV continues to outperform the benchmark, with advantages of the CSV specification for both time horizons considered. The time profile of the log score differences indicate that both, short- and long term rates share a similar LPS pattern. That is, strong gains in accuracy in the crisis and minor improvements prior- and after the crisis. Note that this holds true for both maturities of interest rates considered, where for long term rates, the heteroscedastic variant of the GVAR tends to perform well even in the crisis. This is due to the fact that in financial economics it is typically assumed that those quantities tend to follow random walk processes, which indicates that a prior that centers the system on a random walk is the best choice to forecast interest rates (see Fama, 1990, for a prominent contribution outlining the difficulty to forecast interest rates at short time horizons).

Exchange rates (rer) are, again, more accurately forecasted using the CSV specification. Again, the time profile of the log scores indicates that the inclusion of stochastic volatility leads to more robust predictions with respect to economic outliers. Especially for the one-quarter ahead predictive densities, this can be seen quite clearly. In the period between 2008Q3 and 2009Q2, Figure 1c reveals that, while the log score for the benchmark depreciates severely, the CSV specification does not suffer severe losses in terms of predictive accuracy. Thus the pattern observed for GDP and inflation directly carries over to exchange rate forecasting, outlining the prominent role stochastic volatility plays for such variables.

Finally, moving to equity indices (eq) confirms a pattern seen for most quantities considered so far. While at the one quarter ahead time horizon the prediction gains are pronounced and strongly favour a CSV specification, at the one year ahead horizon this premium is smaller. Moreover, the pattern that the predictive ability of the baseline model drops sharply in the crisis while the CSV specification stays somewhat robust can also be observed for equity prices. This is not surprising given the fact that equity indices tend to feature volatility clustering and sudden changes in the level of the volatility.

Comparing the differences between the one- and four-steps time horizon reveals that when used to conduct short-term forecasts, the CSV specification provides increases in predictive accuracy which are substantial for all variables under consideration. However, and this corroborates the findings in Carriero et al. (2012), for one-year ahead forecasts the CSV specification loses some of its edge. This is due to the fact that for longer forecasting horizons, the predicted volatilities approach their long-run mean, implying that the differences in conditional volatilities between the homoscedastic Minnesota-GVAR and the CSV specification disappears.

[Figure 1 about here.]

[Figure 2 about here.]

4.3 Taking a look at the second moment

The previous subsection highlighted the strong, positive effect of stochastic volatility on density forecasting accuracy. The reason why increases in log scores are common for most variables is mainly due to the fact that stochastic volatility allows for another dimension of flexibility, namely accounting for structural changes in the volatility of the underlying time series at the local level. However, because the GVAR framework also allows us to exploit the cross section, one further explanation of this accuracy-premium could be due to information originating from other countries. For example, our framework implicitly allows for contagion effects in terms of rising macroeconomic volatility, which implies that cross-sectional information is exploited efficiently. This is achieved by noting that other countries influence domestic volatility through the inclusion of weakly exogenous factors.

[Figure 3 about here.]

As a consequence, our model succeeds in exploiting this information and translates this into stronger forecasting performance. It is evident that some countries tend to react faster to economic shocks then other countries. In terms of forecasting this would imply that if country j experiences a sudden rise in volatility at time $t + h$, country i would also be influenced at time $t + h$ through the weakly exogenous factors. This would contemporaneously affect the predictive densities in country i, leading to wider prediction intervals.

Figure 3 depicts the posterior mean of the average country specific volatilities across country groups for selected economies over the hold-out period. Note that all regions experienced sharp increases in volatility between 2007 and 2009. In the recovery following the financial crisis, volatility went back to pre-crisis levels. Note that the recent Euro crisis had severe consequences for the European economies, leading to further increases in the level of volatility. However, from the second half of 2012 onwards volatility levels in Europe are converging towards their pre-2008 levels.

These findings provide some insights on why the GVAR-CSV performs well in in the financial crisis. As described above, while the variance of the predictive densities associated with the CSV specifications increase, the homoscedastic GVAR fails to incorporate the fact that there has been a shift in the underlying volatility. This leads to prediction intervals which are too narrow, ultimately failing to capture sharp movements in the underlying macroeconomic aggregates.

5 Conclusion and Further Remarks

This paper has shown that adding stochastic volatility to the GVAR modeling framework tends to improve the accuracy of density forecasts by large margins. Furthermore, as expected, stochastic volatility specifications tend to produce more robust predictions with respect to the underlying forecasting period. Our GVAR with common stochastic volatility improves upon a linear GVAR coupled with a Minnesota prior for all variables under scrutiny at both time horizons. For all variables under consideration, the CSV specification outperforms its benchmark by quite large margins, especially at the critical one-step ahead time horizon. Moreover, allowing for stochastic volatility helps to make the predictions more robust with respect to the changing magnitudes of economic shocks. This is especially visible in the time period covering the recent financial crisis, where the outperformance of the GVAR-CSV is especially pronounced. This pattern can be seen for most variables under scrutiny, with the exception of interest rates. As a possible avenue of further research the inclusion of time varying coefficients could also prove to have a positive influence on the accuracy of point and density forecasts.

References

Adolfson, Malin, Jesper Lindé, and Mattias Villani, "Forecasting performance of an open economy DSGE model," *Econometric Reviews*, 2007, *26* (2-4), 289–328.

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin, "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 2010, *25* (1), 71–92.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin, "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 2010, *25* (1), 71–92.

Carriero, Andrea, Todd Clark, and Massimiliano Marcellino, "Common drifting volatility in large Bayesian VARs," 2012.

Cesa-Bianchi, Ambrogio, M Hashem Pesaran, and Alessandro Rebucci, "Uncertainty and economic activity: A global perspective," Technical Report, CESifo Working Paper 2014.

Clark, Todd E, "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business & Economic Statistics*, 2011, *29* (3).

Cuaresma, Jesús Crespo, Martin Feldkircher, and Florian Huber, "Forecasting with Bayesian Global Vector Autoregressive Models: A Comparison of Priors," Working Papers 189, Oesterreichische Nationalbank (Austrian Central Bank) March 2014.

Dees, Stephane, Filippo di Mauro, Hashem M. Pesaran, and L. Vanessa Smith, "Exploring the international linkages of the euro area: a global VAR analysis," *Journal of Applied Econometrics*, 2007, *22* (1).

Eisenstat, Eric and Rodney W Strachan, "Modelling Inflation Volatility," Technical Report, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University 2014.

Fama, Eugene F, "Term-structure forecasts of interest rates, inflation and real returns," *Journal of Monetary Economics*, 1990, *25* (1), 59–76.

Feldkircher, M., "A Global Macro Model for Emerging Europe," *Journal of Comparative Economics*, 2014, *forthcoming.*

Frühwirth-Schnatter, Sylvia and Helga Wagner, "Stochastic model specification search for Gaussian and partial non-Gaussian state space models," *Journal of Econometrics*, 2010, *154* (1), 85–100.

Giannone, Domenico, Michele Lenza, and Giorgio E Primiceri, "Prior selection for vector autoregressions," Technical Report, National Bureau of Economic Research 2012.

Giordani, Paolo and Mattias Villani, "Forecasting macroeconomic time series with locally adaptive signal extraction," *International Journal of Forecasting*, 2010, *26* (2), 312–325.

Greenwood-Nimmo, Matthew, Viet Hoang Nguyen, and Yongcheol Shin, "Probabilistic Forecasting of Output Growth, Inflation and the Balance of Trade in a GVAR Framework," *Journal of Applied Econometrics*, 2012, *27*, 554–573.

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

Karlsson, Sune, "Forecasting with Bayesian Vector Autoregressions," *Örebro University Working Paper*, 2012, *12/2012.*

Kastner, Gregor and Sylvia Frühwirth-Schnatter, "Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models," *Computational Statistics & Data Analysis*, 2013.

Koop, Gary and Dimitris Korobilis, "Large time-varying parameter VARs," *Journal of Econometrics*, 2013, *177* (2), 185–198.

Koop, Gary M, "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 2013, *28* (2), 177–203.

Pesaran, M. Hashem, Til Schuermann, and L. Vanessa Smith, "What if the UK or Sweden had joined the euro in 1999? An empirical evaluation using a Global VAR," *International Journal of Finance and Economics*, 2007, *12* (1), 55–87.

—, —, and —, "Forecasting economic and financial variables with global VARs," *International Journal of Forecasting*, 2009, *25* (4), 642–675.

—, —, and S. M. Weiner, "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model," *Journal of Business and Economic Statistics, American Statistical Association*, 2004, *22*, 129–162.

Theil, Henry and Arthur S Goldberger, "On pure and mixed statistical estimation in economics," *International Economic Review*, 1961, *2* (1), 65–78.

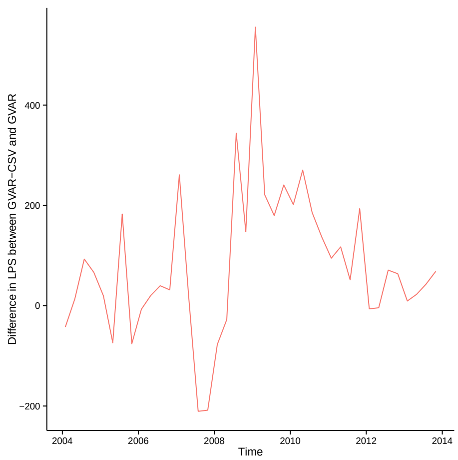
Table 1: Differences in log predictive scores: GVAR-CSV versus GVAR

	1-step-ahead	4-steps-ahead
y	80.759	29.923
Δp	54.446	29.264
i_S	8.843	8.172
i_L	10.071	8.455
rer	84.975	49.272
eq	80.361	26.323

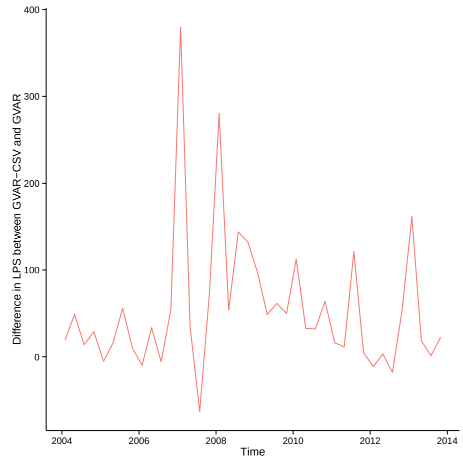
Notes: Average log predictive scores over the hold-out sample.Log scores have been obtained by using the multivariate normal approximation to the predictive density as in Adolfson et al. (2007).

Figure 1: Evolution of Log Predictive Scores over time - 1-step ahead predictive density.

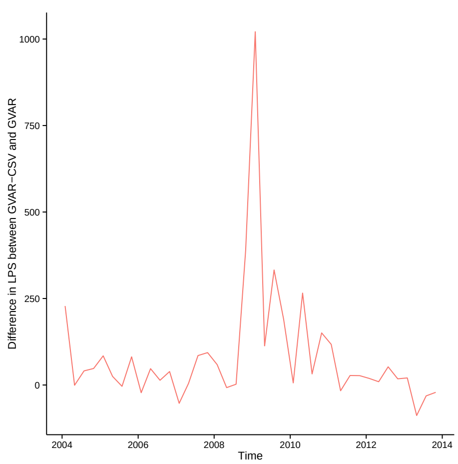
(a) Real GDP



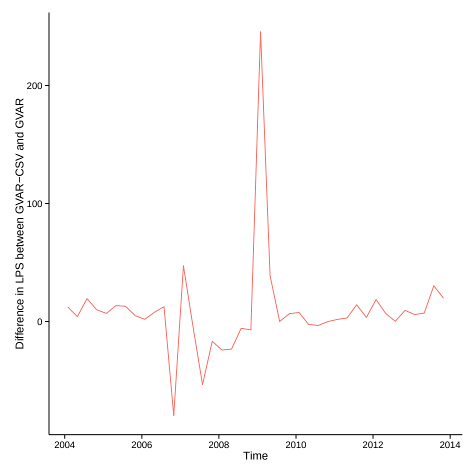
(b) Inflation



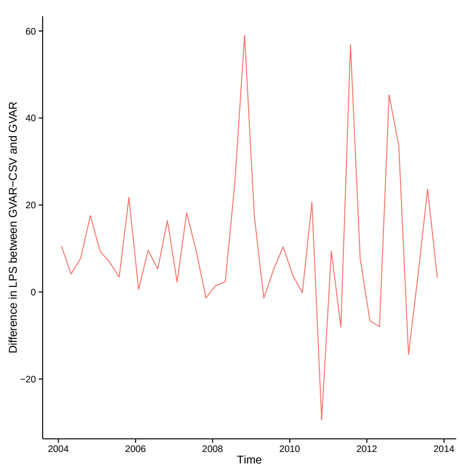
(c) Real Exchange Rate



(d) Short-term interest rates



(e) Long-term interest rates



(f) Equities

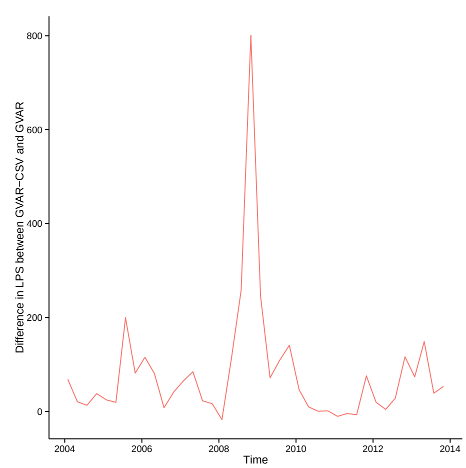
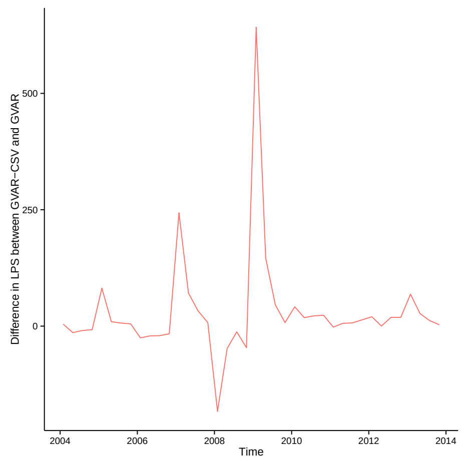
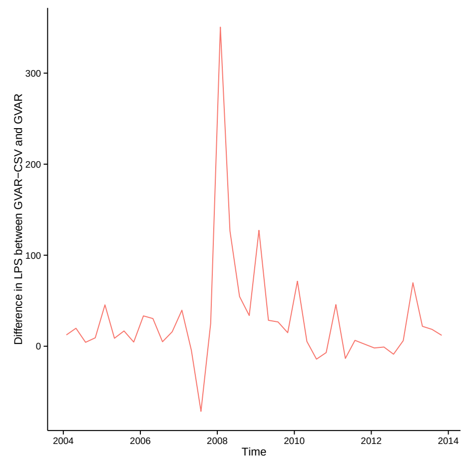


Figure 2: Evolution of Log Predictive Scores over time - 4-steps ahead predictive density.

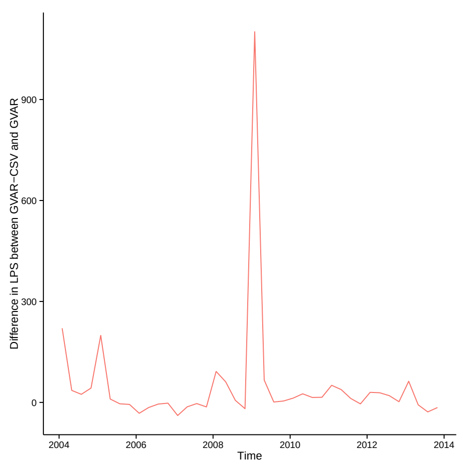
(a) Real GDP



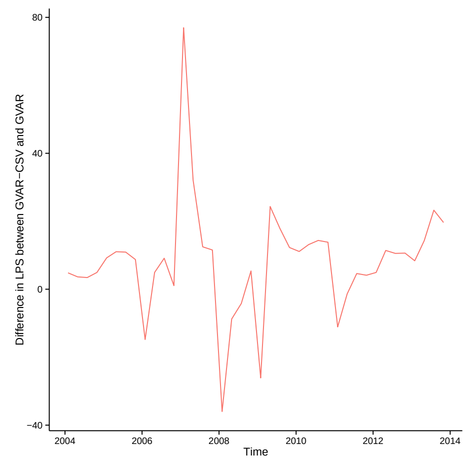
(b) Inflation



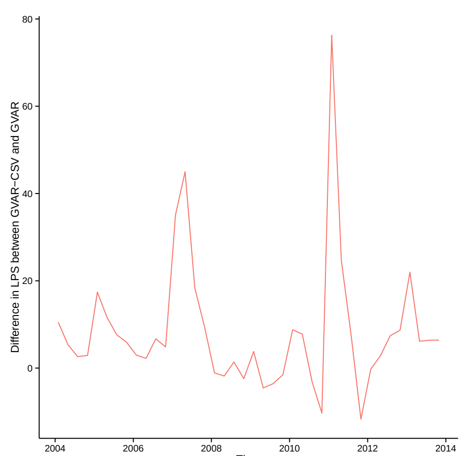
(c) Real Exchange Rate



(d) Short-term interest rates



(e) Long-term interest rates



(f) Equities

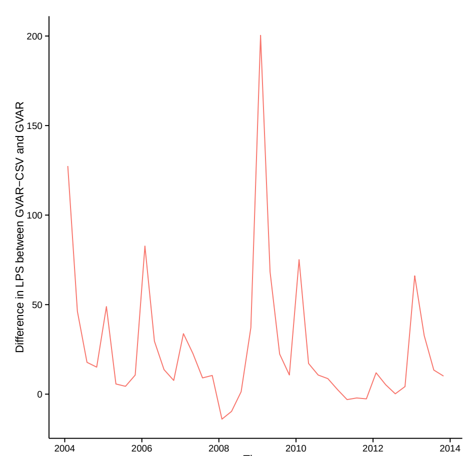
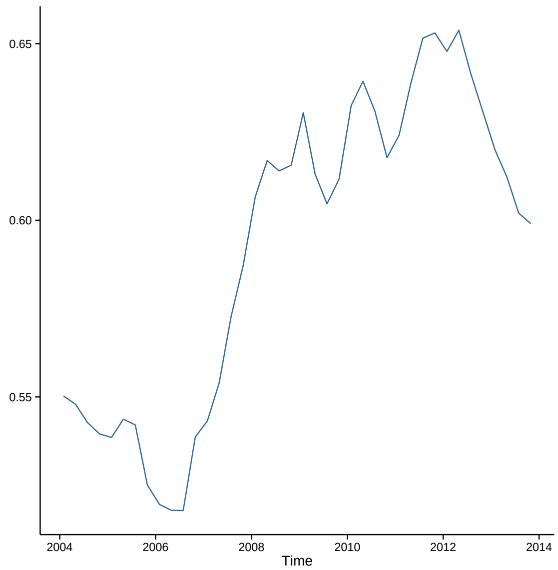
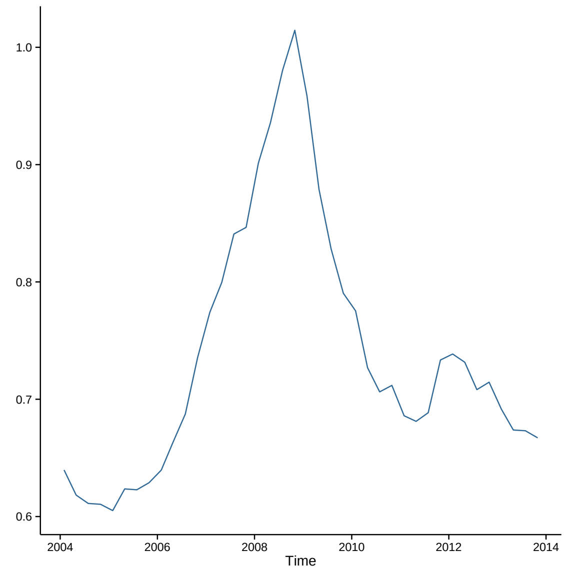


Figure 3: Posterior Mean of volatility factors across countries

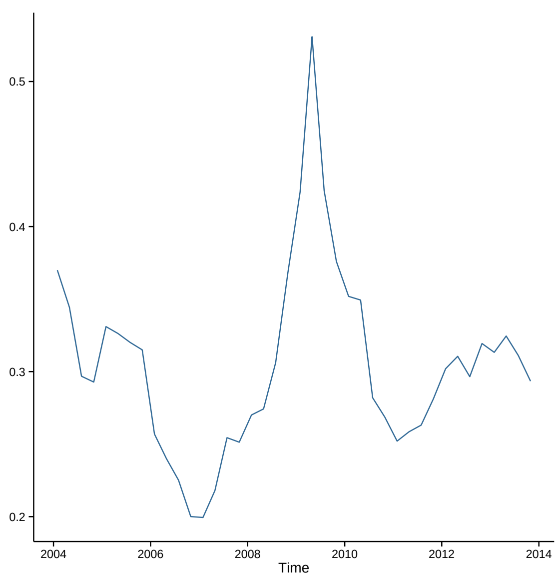
(a) Europe



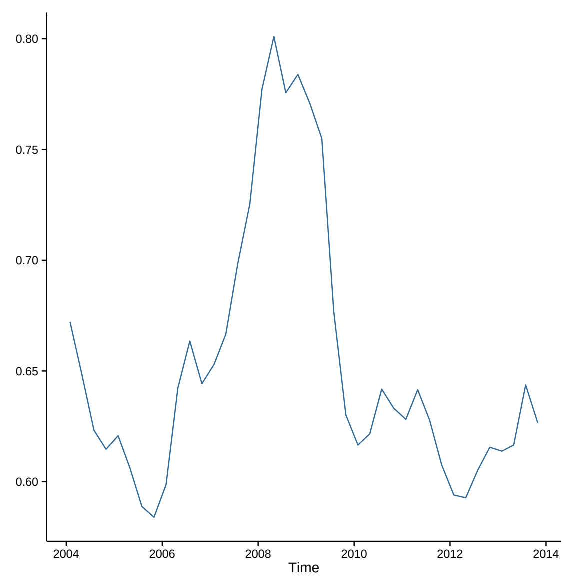
(b) Asia



(c) Latin America



(d) Rest of the World



Appendix A Data and Country coverage

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 U.S. 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 (2014). Source: Crespo Cuaresma et al. (2014)

Table A.2: Country coverage of GVAR model

Europe	Other Developed	Emerging Asia	Latin America	Mid-East and Africa
Austria (AT)	Australia (AU)	China (CH)	Argentina (AR)	Turkey (TR)
Belgium (BE)	Canada (CA)	India (IN)	Brazil (BR)	Saudi Arabia (SA)
Germany (DE)	Japan (JP)	Indonesia (ID)	Chile (CL)	South Africa (ZA)
Spain (ES)	New Zealand (NZ)	Malaysia (MY)	Mexico (MX)	
Finland (FI)	United States (US)	Korea (KR)	Peru (PE)	
France (FR)		Philippines (PH)		
Greece (GR)		Singapore (SG)		
Italy (IT)		Thailand (TH)		
Netherlands (NL)				
Portugal (PT)				
Denmark (DK)				
Great Britain (GB)				
Switzerland (CH)				
Norway (NO)				
Sweden (SE)				

Notes: ISO-2 country codes in brackets. Source: Crespo Cuaresma et al. (2014)