Common Drifting Volatility in Large Bayesian VARs

Common Drifting Volatility in Large Bayesian VARs*

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

Todd E. Clark
Federal Reserve Bank of Cleveland
todd.clark@clev.frb.org

Massimiliano Marcellino
Bocconi University, IGIER and CEPR
massimiliano.marcellino@unibocconi.it

Abstract

The general pattern of estimated volatilities of macroeconomic and financial variables is often broadly similar. We propose two models in which conditional volatilities feature comovement and study them using U.S. macroeconomic data. The first model specifies the conditional volatilities as driven by a single common unobserved factor, plus an idiosyncratic component. We label this model BVAR with General Factor Stochastic Volatility (BVAR-GFSV) and we show that the loss in terms of marginal likelihood from assuming a common factor for volatility is moderate. The second model, which we label BVAR with Common Stochastic Volatility (BVAR-CSV), is a special case of the BVAR-GFSV in which the idiosyncratic component is eliminated and the loadings to the factor are set to 1 for all the conditional volatilities. Such restrictions permit a convenient Kronecker structure for the posterior variance of the VAR coefficients, which in turn permits estimating the model even with large datasets. While perhaps misspecified, the BVAR-CSV model is strongly supported by the data when compared against standard homoskedastic BVARs, and it can produce relatively good point and density forecasts by taking advantage of the information contained in large datasets.

Keywords: stochastic volatility, prediction, forecasting

J.E.L. Classification: C11, C13, C33, C53.

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. We would like to thank the Editor, two anonymous referees, Marco Del Negro, Ulrich Müller, Haroon Mumtaz, Giorgio Primiceri, Frank Schorfheide, Mark Watson, Jonathan Wright, participants at the 7th ECB Forecasting Conference, 2012 NBER Summer Institute, RCEA Bayesian Econometric Workshop, and a seminar at Rotterdam University for helpful comments on a previous draft. Carriero gratefully acknowledges support for this work from the Economic and Social Research Council [ES/K010611/1].

1 Introduction

Several recent papers have shown that the use of large vector autoregressions (VARs) produces significant improvement in both forecasting and structural analysis of macroeconomic data. In particular, contributions such as Banbura, Giannone, and Reichlin (2010), Carriero, Clark, and Marcellino (2015), Giannone, Lenza, and Primiceri (2012) and Koop (2013) all point out that a system of 15-20 variables performs better than smaller systems in point forecasting and structural analysis. All these contributions exploit a natural conjugate prior for the VAR coefficients which yields posterior distributions featuring a particularly convenient Kronecker structure that makes estimation fast.

Carriero, Clark, and Marcellino (2015), Giannone, Lenza, and Primiceri (2012) and Koop (2013) also consider density forecasting, and again conclude that the use of a large information set is beneficial. However, the VARs considered in such studies feature homoskedastic disturbances, while there is convincing evidence that instead macroeconomic fluctuations are characterized by time-varying volatilities (e.g., Clark (2011), Clark and Ravazzolo (2014), Cogley and Sargent (2005), D'Agostino, Gambetti and Giannone (2013), and Primiceri (2005)). To make this point more apparent, Figure 1 plots the estimated volatilities of 14 major US macroeconomic time-series, obtained using univariate AR models with stochastic volatility, and shows clear evidence of strong time variation in the volatilities.

While the assumption of homoskedastic disturbances can be considered mild when the interest is limited to point forecasts, such an assumption appears to be strong for density forecasting. However, relaxing the assumption of homoskedasticity of the VAR disturbances poses problems if the dimension of the VAR is large, because the introduction of drifting volatilities leads to the loss of symmetry in the model, which in turn implies that estimation of the system becomes rapidly unmanageable. Homoskedastic VAR models are SUR models featuring the same set of regressors in each equation. This symmetry across equations means that homoskedastic VAR models have a Kronecker structure in the likelihood, and can therefore be estimated via OLS equation by equa-

tion. In a Bayesian setting the symmetry in the likelihood transfers to the posterior, as long as the prior used also features a Kronecker structure.

Equation-specific stochastic volatility breaks this symmetry because each equation is driven by a different volatility. The challenge with such a model is that drawing the VAR coefficients from the conditional posterior involves computing a (variance) matrix with the number of rows and columns equal to the number of variables squared times the number of lags (plus one if a constant is included). The size of this matrix increases with the square of the number of variables in the model, making CPU time requirements highly nonlinear in the number of variables. In light of these challenges, the studies of VARs with stochastic volatility by Clark (2011), Clark and Ravazzolo (2014), Cogley and Sargent (2005), D'Agostino, Gambetti and Giannone (2013), and Primiceri (2005) have been limited to a handful of variables (3 to 5). In one exception, Koop and Korobilis (2013) propose a computational (not fully Bayesian) shortcut that allows for time-varying volatility (roughly speaking, using a form of exponential smoothing of volatility) in a large VAR, and show that their model improves the accuracy of point and density forecasts.

In this paper we propose a (fully Bayesian and) computationally tractable way to introduce stochastic volatility in a large VAR. Our method is based on the observation that while of course different variables have different volatilities, a common pattern can be observed. For example, looking again at the graphs in Figure 1, it is apparent that U.S. macroeconomic variables show a common pattern of higher volatility in the 1970s, a marked decrease in volatility starting in the early 1980s — the Great Moderation — and a new increase with the onset of the financial crisis. This commonality is in turn reflected in the fact that most of the total variation in the volatilities of macro variables is well summarized by the first principal component. For example, for the 14 variables depicted in Figure 1, the first principal component estimate explains about 70% of the variation in the individual volatility time series.

We propose two models featuring comovement in the volatilities. The first model specifies the conditional volatilities as driven by a single common unobserved factor, plus an idiosyncratic

component. We label this model BVAR with General Factor Stochastic Volatility (BVAR-GFSV). Estimating this model using U.S. quarterly data from 1965 to 2013, we find that the loss in marginal likelihood from imposing a common factor on the volatilities is moderate, and that it provides economically sensible estimates of the time varying variance of macroeconomic variables. These findings point towards the conclusion that there is significant comovement in macroeconomic volatilities.

While the BVAR-GFSV offers a good description of the data, it cannot easily be estimated when the dimension of the cross-section increases above a handful of variables. Therefore, we propose a second model which is nested in the BVAR-GFSV and is computationally manageable for large datasets such as the typical macroeconomic dataset of 15-20 variables (or even more). This model, which we label BVAR with Common Stochastic Volatility (BVAR-CSV), imposes two restrictions on the equation relating the volatilities to the common factor, namely that i) there is no idiosyncratic component for the conditional volatilities, and ii) all the conditional volatilities have a factor loading of 1, which implies that the order of magnitude of the movements in volatility is proportional across variables. The use of the single common factor and of these two restrictions allows us to exploit a Kronecker structure of the conditional posterior densities, thereby allowing us to handle large models.

We show that, as gauged by the marginal likelihood, the restrictions implied by the BVAR-CSV are rejected when compared to the more general BVAR-GFSV specification, so that the former model is likely to be misspecified. However, the BVAR-CSV has the important advantage that it can be estimated for large datasets. Moreover, the empirical evidence we provide is strongly in favor of the BVAR-CSV when it is compared against the standard homoskedastic BVAR, which has been used in several empirical analyses to study medium and large datasets. A forecasting exercise based on real-time data confirms this finding. In both a small model and a large model, our proposed BVAR-CSV specification performs relatively well in both point and density forecasting.

Our proposed volatility models treat the commonality as multiplicative. In the BVAR-CSV

implementation we need both the single factor and the multiplicative structure in order to be able to define a prior and factor out volatility in such a way as to exploit the Kronecker structure that is needed to speed up the VAR computations. Prior work by Pajor (2006) considered the same basic model of volatility for the errors of a VAR(1) process, in just a few variables, without the VAR prior we incorporate to speed up computations. Osiewalski and Pajor (2009) and references therein have considered common volatility within GARCH-type specifications.

Some other papers introduce the commonality in volatility as additive. For example, in an asset return context, Chib, Nardari, and Shephard (2002, 2006) and Jacquier, Polson and Rossi (1995) employ a factor structure multivariate stochastic volatility model. In a macro context, in a setup similar to that used in some finance research, Del Negro and Otrok (2008), Liu, Mumtaz and Theophilopoulou (2014), and Mumtaz and Surico (2012) develop a factor model with stochastic volatility. Viewed this way, the factor structure multivariate stochastic volatility model or factor model with stochastic volatility is somewhat different from the one proposed here: in the BVAR-GFSV and BVAR-CSV we have a VAR that captures cross-variable correlations in conditional means and captures a common factor in just volatility; in these other models, the factor captures both cross-variable correlations in conditional means and drives commonality in volatility.

To establish the value of our proposed model(s), we compare volatility estimates, measures of in-sample fit, and forecast accuracy (both point and density) in both small and large VARs estimated with and without variation in the volatilities. We find that common stochastic volatility in general improves forecast accuracy, and we confirm the finding that the use of a large cross-section helps. We interpret these results as evidence that the large BVAR-CSV model efficiently summarizes the information in a large dataset and successfully accounts for changing volatility, outperforming the conventional approach that treats the volatility of each variable as constant.

The structure of the paper is as follows. Section 2 presents the models, discusses the priors, and derives the posteriors (with additional details in the Appendix). Section 3 describes the MCMC implementation. Section 4 presents our US-based evidence, including full-sample volatility estimates

and our forecasting exercise. Section 5 summarizes the main results and concludes.

2 Modeling comovement in volatilities

2.1 BVAR with General Factor Stochastic Volatility (BVAR-GFSV)

Let y_t denote the $n \times 1$ vector of model variables and p the number of lags. Define the following: Π_0 = an $n \times 1$ vector of intercepts; $\Pi(L) = \Pi_1 - \Pi_2 L - \cdots - \Pi_p L^{p-1}$, with each Π_i an $n \times n$ matrix, $i = 1, \ldots, p$; and A = an $n \times n$ lower triangular matrix with ones on the diagonal.

The BVAR-GFSV model takes the form:

$$y_t = \Pi_0 + \Pi(L) y_{t-1} + v_t, \tag{1}$$

$$v_t = \quad _t \varepsilon_t, \ \varepsilon_t \sim iid\ N(0, I), \tag{2}$$

which implies a time varying variance for the disturbances:

$$Var(v_t) = \Sigma_t = A^{-1} \Lambda_t A^{-1'} \ . \tag{3}$$

Here Λ_t is a diagonal matrix with generic j-th element:

$$\lambda_{jt} = f_t^{\beta_j} \cdot h_{i,t}, \tag{4}$$

so that the log-volatilities follow a linear factor model:

$$\ln \lambda_{jt} = \beta_j \ln f_t + \ln h_{j,t}, \tag{5}$$

where f_t is a common factor and $h_{j,t}$ is the idiosyncratic component associated with the j-th variable in the *VAR*. The diagonality of the matrix Λ_t implies that the generic j-th element of the rescaled *VAR* disturbances $\tilde{v}_t = Av_t$ is given by $\tilde{v}_{j,t} = \lambda_{jt} \varepsilon_{jt}$. Taking logs of squares of $\tilde{v}_{j,t}$ yields the following set of observation equations:

$$\ln \tilde{v}_{j,t}^2 = \beta_j \ln f_t + \ln h_{j,t} + \ln \varepsilon_{j,t}^2, \ j = 1, \ldots, n. \tag{6}$$

The model is completed by specifying laws of motion for the unobserved states:

$$\ln h_{j,t} = \gamma_{j,0} + \gamma_{j,1} \ln h_{j,t-1} + e_{j,t}, \quad j = 1, \ldots, n, \tag{7}$$

$$\ln f_t = \psi \ln f_{t-1} + u_t. \tag{8}$$

For identification, we set $\beta_1 = 1$ and assume $\ln f_t$ to have zero unconditional mean. We further assume that the innovations to volatilities are jointly distributed as $i.i.d.\ N(0, \Phi)$ and independent among themselves, so that

$$Var(e_{1,t}, \ldots, e_{n,t}, u_t) = \Phi = diag(\varphi_1, \ldots, \varphi_n, \varphi_{n+1}). \tag{9}$$

Equations (6)-(8) form a state-space system with unobserved states $\alpha_t = (\ln h_{1,t}, \ln h_{2,t}, \ldots, \ln h_{n,t}, \ln f_t)'$, which can be handled using the Kim, Shephard, and Chib (1998) sampler.

The model described above is related to Cogley and Sargent (2005) and Primiceri (2005). In particular, both these papers assume that there is no factor structure in the volatilities, which amounts to setting $\beta_j = 0$. However, Primiceri's (2005) model is more general in that it permits the innovations to the volatilities to be correlated across variables, while in our specification they are not, and any correlation among volatilities are forced onto the common factor, a restriction that is standard in factor model analysis. For comparison, in our empirical analysis we will also consider a BVAR with the stochastic volatility specification of Primiceri (2005), as described below in Section 2.3.

2.2 BVAR with Common Stochastic Volatility (BVAR-CSV)

The model described in equations (1)-(9) is very general, but it cannot be easily estimated for systems with more than a handful of variables. This happens because the posterior variance of the VAR's conditional mean coefficients has dimension $(np + 1)n$. However, it is possible to design a model in which this matrix — while remaining of the same dimension — can be conveniently factorized. The way to achieve this is to impose some restrictions on the BVAR-GFSV and introduce

a specific prior for the VAR coefficients.

The needed restrictions make the factor loadings β_j equal to 1 for all variables and make idiosyncratic components $h_{j,t}$ constants for all time. In this case, the constants h_j would capture differences in variance scales for each variable (the matrix A of the general model is defined to have values of 1 for all diagonal elements). We will instead, but equivalently, redefine A to have a non-unity diagonal, and use this matrix, denoted $\tilde{A}$, to capture differences in variance scales, and make the idiosyncratic components $h_{j,t}$ equal 1 for all time. As becomes clear below, this will allow us to replace a slightly more complicated step for sampling the rows of A from normal distributions with a step for sampling a constant variance matrix Σ from an inverted Wishart distribution.

More specifically, we impose the following restrictions on equation (4):

$$\beta_j = 1, \; h_{j,t} = 1, \tag{10}$$

for all $t = 1, \ldots, T$ and $j = 1, \ldots, n$. This implies that the generic element of the matrix Λ_t is given by $\lambda_{jt} = f_t$, and therefore $\Lambda_t = f_t I_n$. Thus, equation (2) reduces to:

$$v_t = \tilde{A}^{-1} \varepsilon_t, \; \varepsilon_t \sim iid\, N(0, I), \tag{11}$$

where $\tilde{A}$ is a lower triangular matrix and f_t is a scalar process. By defining $\Sigma = \tilde{A}^{-1} \tilde{A}^{-1'}$, equation

(3) simplifies to:

$$Var(v_t) = \Sigma_t = f_t \Sigma. \tag{12}$$

The fact that the time varying variance Σ_t can be decomposed into the product of a time varying *scalar* and a constant scaling matrix Σ implies that the likelihood function of the VAR features a Kronecker structure, and therefore a posterior with a Kronecker structure can be easily obtained by simply choosing an appropriate conjugate prior. Under these restrictions, the state-space system describing the evolution of the rescaled reduced form squared errors $\tilde{v}^2_{j,t}$ reduces to:

$$\ln \tilde{v}^2_{j,t} = \ln f_t + \ln \varepsilon^2_{j,t}, \; j = 1, \ldots, n, \tag{13}$$

$$\ln f_t = \psi \ln f_{t-1} + u_t, \; Var(u_t) = \varphi, \tag{14}$$

where ψ and φ are scalars.

The key advantage of imposing the restrictions in (10) is that it allows us to use the BVAR-CSV model to estimate larger systems. While the posterior variance of the VAR coefficients is of course still of dimension $(np + 1)n$, its manipulation will involve $(np + 1)^3 + n^3$ elementary operations, rather than $((np + 1)n)^3$. To give an example, in a system of 20 variables and 4 lags, the manipulation of this matrix would involve 4251528000 elementary operations, but when the matrix has a Kronecker structure, the number of operations reduces to just 539441. Such gains become larger as the cross-sectional dimension or the number of lags in the VAR increases. As noted above, being able to estimate a model with a large dataset and rich dynamics is important, as there is convincing evidence that a large information set is useful in forecasting and structural analysis.

Of course, the possibility of estimating a large system does not come without a cost. In particular, the restrictions in (10) do not necessarily hold in a typical dataset of macroeconomic and financial variables. Restriction (10) implies that when the volatility factor f_t increases/decreases, the order of magnitude of the increase/decrease in volatility is the same for all the variables, so for example the volatility of all variables will double if f_t doubles. Instead, Figure 1 suggests that the relative amount of the increase in volatility for some variables (e.g. the federal funds rate) is larger than for other macroeconomics variables. Restriction (10) implies that there are no idiosyncratic components and the common factor f_t is the only source of variation in the volatilities, which clearly might not be enough to fully describe the pattern of volatilities shown in Figure 1. It can be argued that even without idiosyncratic components, a good description could be obtained by adding more common factors. Such an argument is correct, and indeed we have also experimented with such a model. However the use of more than one factor would still encounter the same computational difficulties as having the idiosyncratic components, thereby excluding the possibility of using a large dataset. Also, the BVAR-CSV is very likely less misspecified than a model that assumes no time variation in volatilities, such as the standard BVAR that has been successfully used

for forecasting with large datasets in many recent contributions mentioned above.

Therefore, there is a trade-off between using the more general BVAR-GFSV model and giving up on using the information available in a large dataset, or using a model like the BVAR-CSV, which may be misspecified, in order to take advantage of large information sets. Drawing on George Box's statement that "all models are wrong, but some are useful," we will provide evidence that the BVAR-CSV specification is useful. In particular, in our empirical application we show that the BVAR-CSV provides (in both small and large models) improvements in density forecasts against the standard homoskedastic BVAR, and (in small models) has a performance comparable to both the BVAR-GFSV and the BVAR with conventional stochastic volatility.

2.3 BVARs with individual stochastic volatilities and constant volatility (BVAR-SV and BVAR)

As noted above, for reference, we will consider estimates from a VAR with individual stochastic volatilities for each variable, along the lines of Cogley and Sargent (2005) and Primiceri (2005). The BVAR-SV model we use can be obtained by imposing on the BVAR-GFSV the restriction $\beta_j = 0$ for all j, and by replacing (9) with:

$$Var(e_{1,t}, \ldots, e_{n,t}) = \Phi, \tag{15}$$

where Φ is a full $n \times n$ matrix that is not restricted to be diagonal. Departing from Cogley and Sar- gent (2005) and Primiceri (2005), we treat the volatilities as following the autoregressive processes in (7), rather than random walks, which may have consequences on the estimated volatility paths. Clark and Ravazzolo (2014) find that AR and random walk specifications perform comparably in out-of-sample forecasting.

We will also consider some results from a BVAR with constant volatility, of the form:

$$y_t = \Pi_0 + \Pi(L)y_{t-1} + v_t, \; v_t \sim N(0, \Sigma). \tag{16}$$

For this model, we follow other work on large VARs (e.g., Banbura, Giannone, and Reichlin (2010) and Carriero, Kapetanios, and Marcellino (2009)) and use the Normal-inverted Wishart prior and posterior detailed in such studies as Kadiyala and Karlsson (1997).

2.4 Priors for BVAR-CSV

As the focus of this paper is on the use of large datasets, we focus on the description of the BVAR-CSV, which can be estimated with a large cross-section of data. A detailed description of the priors and posteriors for the BVAR-GFSV and priors for the other models can be found in the Appendix.

The parameters of the model consist of the following: the parameters in the conditional mean matrices Π_0 and $\Pi(L)$, which we collect in a $k \times n$ (where $k = 1 + np$) matrix $\Pi = (\Pi_0, \Pi_1, \ldots, \Pi_p)'$; the elements of Σ; and the coefficients ψ and φ appearing in the state space system (13)-(14). The model also includes the latent states f_t, $t = 1, \ldots, T$, which we collect in the vector $\mathbf{f}$. We use $N(a, b)$ to denote a normal distribution (either univariate or multivariate) with mean a and variance b, $IW(a, b)$ to denote an inverse Wishart distribution with scale matrix a and degrees of freedom b, and $IG(a, b)$ to denote an inverse gamma distribution with scale term a and degrees of freedom b.

We specify the following priors for the parameter blocks of the model:

$$\text{vec}(\Pi)|\Sigma \sim N(\text{vec}(\underline{\mu}_\Pi), \underline{\Omega}_\Pi), \tag{17}$$

$$\Sigma \sim IW(d_s \cdot \underline{\Sigma}, d_s), \tag{18}$$

$$\psi \sim N(\underline{\mu}_\psi, \underline{\Omega}_\psi) \tag{19}$$

$$\varphi \sim IG(d_\varphi \cdot \underline{\varphi}, d_\varphi). \tag{20}$$

A detailed description of the prior moments is given below; here we just emphasize that to make estimation with large models tractable, the prior variance for $\text{vec}(\Pi)$ is specified with a factorization that permits a Kronecker structure. Specifically, we use a prior conditional on Σ, of the following

form:

$$\underline{\Omega}_{\Pi} = \Sigma \otimes \underline{\Omega}_0, \tag{21}$$

where $\underline{\Omega}_0$ incorporates the kind of symmetric coefficient shrinkage typical of the natural conjugate Normal-Wishart prior. Under the usual Minnesota-style specification of the Normal-Wishart prior for $\underline{\Omega}_0$, the prior variance takes account of volatility (and relative volatilities of different variables) by using variance estimates from some training sample. Note that the use of a prior for the coefficients conditional on volatility is in line with the natural conjugate Normal-Wishart prior, but it does depart from the setup of Clark (2011) and Clark and Davig (2011), in which, for a VAR with *n* individual stochastic volatilities, the coefficient prior was unconditional.

The prior used here, combined with the assumption of a single volatility factor, implies that the posterior distribution of the VAR coefficients, conditional on Σ and **f**, will have a variance featuring a Kronecker structure. The computations required to draw from such a distribution via MC sampling are of order $O(n^3 p^3)$ rather than of order $O(n^6 p^3)$. Further detail on how these assumptions yield computational improvements is given in the Appendix.

2.5 Posteriors for BVAR-CSV

The parameters Π, Σ, ψ, and φ have closed-form conditional posterior distributions which we present here. Draws from these conditionals will constitute Gibbs sampling steps in our MCMC algorithm. As detailed below, we will also use a Gibbs sampling to draw the volatilities, with the basic algorithm of Kim, Shephard, and Chib (1998).

The conditional posterior distributions of Π, Σ, ψ, and φ take the following forms:

$$\text{vec}(\Pi)|\Sigma,\mathbf{f},y \sim N(\text{vec}(\bar{\mu}_\Pi),\bar{\Omega}_\Pi), \tag{22}$$

$$\Sigma|\Pi,\mathbf{f},y \sim IW\left(d_s\cdot\underline{\Sigma}+\sum_{t=1}^{T} f_t^{-1}v_t v_t',\ d_s+T\right), \tag{23}$$

$$\psi|\Pi,\Sigma,\varphi,\mathbf{f},y \sim N(\text{vec}(\bar{\mu}_\psi),\bar{\Omega}_\psi), \tag{24}$$

$$\varphi|\Pi,\Sigma,\psi,\mathbf{f},y \sim IG\left(d_\varphi\cdot\underline{\varphi}+\sum_{t=1}^{T} u_t^2,\ d_\varphi+T\right) \tag{25}$$

where y is a nT-dimensional vector containing all the data.

The mean and (inverse) variance of the conditional posterior normal distribution for vec(Π) in (22) take the following forms:

$$\text{vec}(\bar{\mu}_\Pi) = \bar{\Omega}_\Pi\left(\text{vec}\left(\sum_{t=1}^{T} X_t y_t'\right)\Sigma^{-1}+\underline{\Omega}_\Pi^{-1}\text{vec}(\underline{\mu}_\Pi)\right) \tag{26}$$

$$\bar{\Omega}_\Pi = \left(\Sigma^{-1}\otimes\left(\sum_{t=1}^{T} f_t X_t X_t'\right)+\underline{\Omega}_\Pi^{-1}\right)^{-1}. \tag{27}$$

The mean and variance of the conditional posterior normal distribution for ψ in (24) are simpler, because ψ is just a scalar and the relevant process is conditionally homoskedastic:

$$\bar{\mu}_\psi = \bar{\Omega}_\psi\left(\varphi^{-1}\sum_{t=1}^{T} f_{t-1}f_t+\underline{\Omega}_\psi^{-1}\underline{\mu}_\psi\right) \tag{28}$$

$$\bar{\Omega}_\psi = \left(\underline{\Omega}_\psi+\varphi^{-1}\sum_{t=1}^{T} f_{t-1}^2\right)^{-1}. \tag{29}$$

As we have already discussed, the key to the computational advantage of this model is the Kronecker structure of the conditional posterior variance. This Kronecker structure is obtained by using both a single, multiplicative volatility factor and the conditional prior described above. If, instead, one were to use a model in which each variable in the VAR featured an individual stochastic volatility factor, without imposing a factor structure, as has become common with small

VAR specifications, the expression of the variance matrix in (27) would have to be replaced with:

$$\bar{\Omega}_{\Pi}^{-1} = \Omega_{\Pi}^{-1} + \sum_{t=1}^{T} (\Sigma_t^{-1} \otimes X_t X_t'), \tag{30}$$

where Σ_t can not be factorized as $\Sigma_t = f_t \Sigma$. Instead, under the restrictions (10), the matrix Σ_t simplifies to $\Sigma_t = f_t \Sigma$ and the term $\sum_{t=1}^{T} (\Sigma^{-1} \otimes X_t X_t')$ in (30) can be written as $\Sigma^{-1} \otimes \sum_{t=1}^{T} \frac{1}{f_t} X_t X_t'$, which using equation (21) yields the simplified expression for $\bar{\Omega}_{\Pi}^{-1}$ appearing in (27).

The computations required to manipulate the matrix $\bar{\Omega}_{\Pi}^{-1}$ are of order $O(n^3 p^3)$ when using (27), while they are of order $O(n^6 p^3)$ if one were to use (30). In particular, to perform a draw from a normal having (inverse) variance (30), it is necessary to generate a nk-dimensional vector g of standard normal variables and: i) compute $\bar{\Omega}_{\Pi}$ by inverting $\bar{\Omega}_{\Pi}^{-1}$, ii) compute the Cholesky factor of $\bar{\Omega}_{\Pi}$, and iii) multiply the Cholesky factor of $\bar{\Omega}_{\Pi}$ by g. Each of these three steps involves $n^3 k^3$ elementary operations, so given that $k = np + 1$ the time complexity of this procedure is $O(n^6 p^3)$, where in computer science the time complexity of an algorithm quantifies the amount of time taken by an algorithm to run. Instead, to perform a draw from a normal having variance (27), one can draw a $n \times k$ random matrix G of standard normal variables and then use the following formula:

$$\Pi^{draw} = \bar{\mu}_{\Pi} + chol\left[\Omega_0^{-1} + \sum_{t=1}^{T} \left(\frac{1}{f_t} X_t X_t'\right)\right]^{-1} \cdot G \cdot chol[\Sigma]', \tag{31}$$

in which the more demanding computations are only the inversion and Cholesky decomposition of the matrix in the square brackets, which are both of order $O(n^3 p^3)$, as further detailed in the Appendix. Therefore the time complexity implied by (30) is n^3 times larger than that implied by (27), and the use of (30) becomes rapidly unmanageable as the size of the system increases.

Under the *CSV* specification, the expression for the posterior mean of the coefficient matrix appearing in (26) can also be re-written in an equivalent form that may often be more computationally efficient. This equivalent form is obtained by defining data vectors normalized by the standard deviation of volatility, to permit rewriting the *VAR* in terms of conditionally homoskedas-

tic variables. Specifically, let $\tilde{y}_t = f_t^{-0.5} y_t$ and $\tilde{X}_t = f_t^{-0.5} X_t$ and collect these rescaled data for $t = 1, \ldots, T$ in the full-data matrices $\tilde{X}$ and $\tilde{y}$. Then, the posterior mean of the coefficients can be

equivalently written as

$$\bar{\mu}_{\Pi} = \tilde{X}'\tilde{X} + \underline{\Omega}_0^{-1} + \tilde{X}'\tilde{y}, \qquad (32)$$

which can be used directly in (32) thereby further reducing the number of elementary operations required at every step of the MCMC sampler.

2.6 Volatility

Our treatment of volatility draws on Primiceri's (2005) implementation of the Kim, Shephard, and Chib (1998) algorithm (hereafter, KSC algorithm). As indicated above, v_t denotes the reduced form residuals of the VAR and $\tilde{v}_t = \tilde{A}v_t$ are the rescaled residuals. We further define $v^*_{j,t} = \ln(\tilde{v}^2_{j,t} + \bar{c})$, where $\bar{c}$ denotes an offset constant used in the KSC algorithm. With this notation, we can establish the measurement equation of a state-space system with non-Gaussian errors:

$$v^*_{j,t} = \ln f_t + \ln \varrho^2_{j,t}. \qquad (33)$$

The state equation is simply the AR process of the common factor given by (14).

In the equations above $\ln \varrho^2_{j,t}$ is not Gaussian, but $\varrho_{j,t}$ is a Gaussian process with unit variance, and with this setup we can use the mixture of normals approximation of KSC to estimate volatility with a Gibbs sampler, first drawing the states of the mixture and then drawing volatility conditional on the states. Primiceri (2005) details the steps required, which include the Kalman filter (forward) and a (backward) simulation smoother to draw volatility from a conditional distribution that is normal. In the BVAR-CSV, the key difference with respect to Primiceri's (2005) implementation is that there is just one volatility process (state), rather than n processes (states). Instead, the BVAR-GFSV model is more general than Primiceri's (2005) specification in this respect, as it features $n + 1$ volatility processes, although it also imposes a restriction that the idiosyncratic components be uncorrelated, forcing all comovement onto the common factor. We should also note that we depart from these previous studies in using the improved 10-state mixture approximation of Omori,

et al. (2007) instead of KSC's original 7-state approximation, and we incorporate the reordering of Primiceri's steps suggested in Del Negro and Primiceri (2014).

3 Implementation

3.1 Specifics on priors

For our proposed BVAR-CSV model, we set the prior moments of the VAR coefficients along the lines of the common Minnesota prior (see, e.g., the exposition in section 3.2.1 of Karlsson (2013)), without cross-variable shrinkage:

$$\underline{\mu}_{\Pi} = 0, \text{ such that } E[\Pi_l^{(ij)}] = 0 \;\forall\; i, j, l \tag{34}$$

$$\underline{\Omega}_0 \text{ such that the entry corresponding to } \Pi_l^{(ij)} = \begin{cases} \frac{\theta^2}{l^2\sigma_j^2} & for\ l > 0 \\ \varepsilon^2 & for\ l = 0 \end{cases}. \tag{35}$$

With all of the variables of our VAR models transformed for stationarity (we use growth rates of GDP, the price level, etc.), we set the prior mean of all the VAR coefficients to 0. The variance matrix $\underline{\Omega}_0$ is defined to be consistent with the usual Minnesota prior variance, which is a diagonal matrix.

We note that our proposed BVAR-CSV specification can also be directly applied to models in levels with unit root priors, with the appropriate modification of the prior means on the coefficients. Including priors on sums of coefficients and initial observations as in such studies as Sims and Zha (1998) is also possible.

The shrinkage parameter θ measures the tightness of the prior: when $\theta \to 0$ the prior is imposed exactly and the data do not influence the estimates, while as $\theta \to \infty$ the prior becomes loose and results will approach standard *GLS* estimates. Clearly, the choice of the hyperparameter θ is key and several possibilities are available. Giannone, Lenza, and Primiceri (2012) provide a sampler to estimate the hyperparameters of a homoskedastic BVAR with

conjugate prior, which

could be applied here (conditionally on the volatilities). Alternatively, the parameter θ could be chosen by maximizing the marginal likelihood of the model, as in Del Negro and Schorfheide (2004) and Carriero, Kapetanios, and Marcellino (2012). Still other studies, such as Banbura, Giannone, and Reichlin (2010) and Carriero, et al. (2009), have selected hyperparameters using alternative strategies based on past forecasting performance. As the main focus of this paper is on the variation in the volatilities, we simply adopt the common value (e.g., Sims and Zha 1998) of $\theta = 0.2$, in consideration of the evidence in Carriero, Clark, and Marcellino (2015) that, for a dataset of similar type and dimension, such a strategy does not yield significant losses in forecast accuracy.

The term $1/l^2$ determines the rate at which the prior variance decreases with increasing lag length. To set the scale parameters σ_j^2 we follow common practice (see, e.g., Litterman, 1986; Sims and Zha, 1998) and fix them to the variance of the residuals from univariate AR(4) models, computed for the estimation sample. To make the prior on intercepts loose, we set $\varepsilon = 1000$.

For the constant variance matrix Σ, we use a loosely informative prior based on training sample information. Specifically, we set the prior degrees of freedom at $n + 2$ and the prior mean of Σ at OLS estimates of the residual variance matrix obtained by fitting a VAR(1) to a training sample of 40 observations preceding the estimation sample.

For the slope coefficient of the log factor process (ψ) we use a prior mean of 0.9 and a standard deviation of 0.2. For the prior variance of the innovations to the factor process (φ) we use a mean of 0.01 and set the degrees of freedom to 4. Our prior for φ is similar to (although more generous than) the settings used in studies such as Cogley and Sargent (2005). While this prior is fairly informative, our estimates show the posterior to be rather different from the prior, apparently reflecting considerable information in the data.

3.2 MCMC Algorithm

We estimate the BVAR-CSV model with a Gibbs sampler. As detailed in Del Negro and Primiceri (2014), the treatment of stochastic volatility with the mixture distribution approach of Kim, Shephard, and Chib (1998) necessitates a sampler based on two parameter blocks, for (1) time-varying volatility and (2) the mixture states and the set of model parameters.

Let s^T denote the states of the mixture of normals distribution used in the Kim, Shephard, and Chib (1998) algorithm, Θ denote the parameter block containing Π, Σ, ψ, and φ, and recall that y and $\mathbf{f}$ denote the time series of the data and log-states $\ln f_t$, respectively. We omit some details regarding distributions (notation, essentially) in the interest of brevity; Del Negro and Primiceri (2014) provide further information regarding the details of conditioning necessitated by the treatment of volatility. Note that, in the description below, in some cases conditioning on $\mathbf{f}$ makes it unnecessary to explicitly condition on ψ and φ.

The Gibbs sampler draws in turn from the conditionals $p(\mathbf{f} \mid \Theta, s^T, y)$ and $p(\Theta, s^T \mid \mathbf{f}, y)$.

Step 1: Draw from $p(\mathbf{f} \mid \Theta, s^T, y)$ relying on the state space representation described above and the Kalman filter and simulation smoother of Durbin and Koopman (2001). In the filtering, we set the mean and variance for the initial value of $\ln f_t$ at 0 and 0.5, respectively.

Step 2: Draw from $p(\Theta, s^T \mid \mathbf{f}, y)$ relying on the factorization $p(\Theta, s^T \mid \mathbf{f}, y) \propto p(s^T \mid \Theta, \mathbf{f}, y) \cdot p(\Theta \mid \mathbf{f}, y)$, that is by (i) drawing from the marginal posterior of the model parameters $p(\Theta \mid \mathbf{f}, y)$ and (ii) drawing from the conditional posterior of the mixture states $p(s^T \mid \Theta, \mathbf{f}, y)$. The marginal posterior $p(\Theta \mid \mathbf{f}, y)$ is sampled by further breaking the parameter block into pieces and drawing from the distributions of each parameter piece conditional on the other parameter pieces (steps 2a-2d below), while draws from $p(s^T \mid \Theta, \mathbf{f}, y)$ are obtained using steps similar to those described in Primiceri (2005), modified as needed in light of our common factor process (step 2e below). In more detail, the sub-steps used to produce draws from $p(\Theta, s^T \mid \mathbf{f}, y)$ are as follows.

Step 2a: Draw the matrix of VAR coefficients Π conditional on the data, Σ, and $\mathbf{f}$, using the conditional (normal) distribution for the posterior given in equation (22).

Step 2b: Draw Σ conditional on the data, Π and **f**, using the conditional (IW) distribution for the posterior given in (23).

Step 2c: Draw ψ conditional on the data, Π, Σ, φ, and **f**, using the conditional (normal) distribution for the posterior given in (24).

Step 2d: Draw φ conditional on the data, Π, Σ, ψ, and **f**, using the conditional (IG) distribution for the posterior given in (25).

Step 2e: Draw the states of the mixture of normals distribution conditional on the data, **f**, and the parameter block Θ.

In all cases, we obtain forecast distributions by sampling from the posterior distribution. In particular, for each set of draws of parameters, we: (1) simulate volatility time paths over the forecast interval using the AR process of log volatility; (2) draw shocks to each variable over the forecast interval with variances equal to the draw of Σ_{t+h}; and (3) use the VAR structure of the model to obtain paths of each variable. We form point forecasts as means of the draws of simulated forecasts and density forecasts from the simulated distribution of forecasts. Conditional on the model, the posterior distribution reflects uncertainty stemming from latent states, parameters, and shocks over the forecast interval.

4 Empirical results

4.1 Data and design of the forecast exercise

In our application we focus on two datasets of 4 and 14 variables, at the quarterly frequency. The smaller dataset includes GDP ($\Delta \ln$), the unemployment rate, inflation in the GDP price index ($\Delta \ln$), and the federal funds rate. The larger dataset contains the same variables plus consumption expenditures (denoted PCE, $\Delta \ln$), business fixed investment (denoted BFI, $\Delta \ln$), residential investment ($\Delta \ln$), industrial production ($\Delta \ln$), capacity utilization in manufacturing, payroll employment ($\Delta \ln$), aggregate hours worked (by production and non-supervisory workers in the private sector,

$\Delta \ln$), inflation in the PCE price index ($\Delta \ln$), the 10 year $-$ 3m Treasury bond-bill yield spread, and real stock prices (S&P500 index/PCE price index, $\Delta \ln$).

For the purpose of assessing the convergence and mixing properties of our MCMC algorithm and model fit, we consider full-sample model estimates based on current vintage data taken from the FAME database of the Federal Reserve Board. All growth and inflation rates are measured as annualized log changes (from $t - 1$ to t).

In the real-time forecast analysis, output is measured as GDP or GNP, depending on data vintage. Inflation is measured with the GDP or GNP deflator or price index. For simplicity, hereafter "GDP" and "GDP price index" refer to the output or price series, even though the measures are based on GNP and a fixed weight deflator for some of the sample. Real-time data on GDP, PCE, BFI, residential investment, industrial production, capacity utilization, payroll employment, aggregate hours worked, the GDP price index, and the PCE price index are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM). In the case of unemployment, interest rates, and the (nominal) stock price index, for which real-time revisions are small to essentially non–existent, we abstract from real-time aspects of the data and use current vintage data.

Our analysis of real-time forecasts uses real-time data vintages from 1985:Q1 through 2014:Q2. As described in Croushore and Stark (2001), the vintages of the RTDSM are dated to reflect the information available around the middle of each quarter. For each forecast origin t starting with 1985:Q1, we use the real-time data vintage t containing data through $t - 1$ to estimate the forecast models and construct forecasts for periods t and beyond. The forecast evaluation period runs from 1985:Q1 through 2013:Q4, and the forecasting models are estimated using data starting in 1965:Q1. We report results for forecasts at horizons of 1, 2, 4, 8, and 12 quarters ahead.

As discussed in such sources as Croushore (2006), Romer and Romer (2000), and Sims (2002), evaluating the accuracy of real-time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors. The GDP data available today for, say, 1985, represent

the best available estimates of output in 1985. However, output as defined and measured today is quite different from output as defined and measured years ago. For example, in the mid-1990s, the measure of national output switched from fixed-weight GNP to chain-weighted GDP. Forecasters in 1985 could not have foreseen such changes and the potential impact on measured output. Accordingly, we follow studies such as Clark (2011), Faust and Wright (2009), and Romer and Romer (2000) and use the second available estimates of the real-time measured variables as actuals in evaluating forecast accuracy. For unemployment, interest rates, and the nominal stock price index, the real-time data correspond to the final vintage data.

One final implementation detail to note is that we include four lags in all of our BVARs. With Bayesian methods that provide shrinkage, many prior studies have used the same approach of setting the lag length in line with the data frequency (e.g., Banbura, Giannone, and Reichlin (2010), Del Negro and Schorfheide (2004), Koop (2013), and Sims (1993)).

4.2 MCMC convergence and efficiency

We begin with documenting the convergence properties of our MCMC algorithm for the BVAR-CSV model. All results in the paper are based on a sample of 10,000 retained draws, obtained by sampling a total of 55,000 draws, discarding the first 5,000, and retaining every 5th draw of the post-burn sample.

Table 1 reports summary statistics of inefficiency factors (IF) for the posterior estimates of all groups of model parameters, and rejection rates of Geweke's (1992) test for equal means (within the single chain of draws). A value of the IFs below 20 is generally taken as indication that the chain has satisfactory mixing properties. As for the convergence diagnostic (last column), within the sets of parameter estimates, the rejection rates should be close to the theoretical size of the test (assuming independence of tests, for simplicity), which we have set at 10%. As is clear from the figures in Table 1, our algorithm shows satisfactory mixing and convergence properties.

4.3 Computational comparisons

As we discussed above, the BVAR-GFSV and BVAR-SV models can become unmanageable as the dimension of the system increases, because the posterior covariance matrix of the reduced-form VAR coefficients does not feature a Kronecker structure and its manipulation involves $(np + 1)^3 n^3$ operations. On the other hand, the BVAR-CSV preserves such a Kronecker structure, and the manipulation of the posterior covariance matrix of the reduced-form VAR coefficients involves only $(np + 1)^3 + n^3$ operations. These numerical differences are relatively small in a system with 4 variables and 4 lags, but become exponentially larger as the number of variables used increases.

Table 2 shows (in the middle column) the computational time necessary to produce 10,000 draws from the posteriors of the BVAR, BVAR-CSV, BVAR-GFSV, and BVAR-SV models, with 4 and 14 variables. These figures, obtained using version 8.3 of RATS on a 3.2 GHz Intel CPU, highlight large differences in the computational burdens of the models (and in checks with some related models, the speed of RATS is comparable to the speed of Matlab). For the BVAR-CSV model with 14 variables, producing 10,000 draws takes about 40 minutes, a modest 8 minutes more than it takes to produce the same number of draws from a BVAR with constant volatility. However, for the BVAR-SV and BVAR-GFSV models with 14 variables, it takes more than 500 minutes to produce 10,000 draws. In practice, the times required for estimation are even greater than these numbers indicate, because (as shown in unreported mixing checks) the estimation algorithms for the more general models have poorer mixing properties. We give in the third column the actual number of draws we use in producing model estimates and forecasts reported below (or would use if we were to try to estimate the BVAR-SV and BVAR-GFSV models with 14 variables), which we set in light of the mixing properties of the estimation algorithm. In particular, for the BVAR-GFSV and BVAR-SV models, we used (or would use were large model estimation practicable) higher thinning intervals and more total draws in order to achieve inefficiency factors of the retained chains of draws that are broadly comparable to those reported for the BVAR-CSV model. In a real-time forecasting exercise that involves repeatedly estimating a model (at each forecast origin),

a roughly estimated time requirement of more than 5000 minutes of CPU per forecast for the BVAR-SV model or 10,000 minutes for the BVAR-GFSV model with 14 variables rules out their use.

4.4 Model comparisons: volatility estimates and model fit

For alternative models of time-varying volatility, one key aspect of performance is the ability to capture time variation in volatility and any commonality that might exist. For models using 4 variables, Figure 2 shows estimates of volatility for each variable obtained with BVAR-GFSV, BVAR-SV, and BVAR-CSV specifications. All three estimates display very similar contours. The BVAR-GFSV and BVAR-SV estimates move very closely together. The BVAR-CSV estimate also moves closely with these estimates, although for some variables it shows a little more movement (e.g., GDP growth) and other variables it shows a little less movement (e.g., federal funds rate).

Figure 3 provides alternative estimates of the commonality in volatility. In the 4 variable case, the top panel of the figure shows that the BVAR-CSV estimate of common volatility is very similar to the BVAR-GFSV estimate of common volatility; the correlation between the two estimates is 0.99. This implies that the additional restrictions required to use the BVAR-CSV specification do not much harm the model's ability to capture a common factor in volatility that matches up to a more general specification. In the 14 variable case, as noted above, the BVAR-GFSV is computationally intractable. As a measure of the ability of the BVAR-CSV to capture commonality in volatility, the bottom panel of the figure compares the common factor estimate from the BVAR-CSV model to the first principal component of volatilities obtained from estimates of AR(2) models with stochastic volatility fitted for each variable. As is standard in factor analysis in macroeconomics, we standardized the individual variables' volatilities before computing the principal component. In this 14 variable case, the commonality in the AR-SV estimates of volatility appears to be strong; the principal component explains 66 percent of the overall variation. The common factor estimate from the BVAR-CSV estimate seems to do an effective job of capturing

commonality in volatilities, in the sense that it moves fairly strongly with the principal component from individual AR-SV estimates, with a correlation of 0.85.

Based on these results, it seems that, at least in applications to standard macroeconomic VARs in US data, our GFSV and CSV specifications can effectively capture time variation in conditional volatilities, including the commonality in the time variation.

Of course, in real time, reliable estimation of volatility may prove to be more challenging, in part because, at the end of the sample, only one-sided filtering is possible, and in part because of data revisions. In results omitted for brevity, we have investigated this issue by comparing time series of (BVAR-CSV) volatility estimates from several different real-time data vintages. These comparisons yielded strong commonality in the volatility estimates for each vintage and volatility estimates that were very similar across vintages.

To further compare the models, we use the marginal likelihood (ML) to assess the extent to which the CSV restrictions of (10) harm model fit. Table 3 displays the log ML for a range of models. In the four variable case, we consider four models: a benchmark homoskedastic BVAR, the BVAR-SV, the BVAR-GFSV, and the BVAR-CSV. In the 14 variable case, in light of the computational constraints described above, we consider just a homoskedastic BVAR and the BVAR-CSV. For the models with time-varying volatilities, we compute the ML using both the Chib (1995) method and the Modified Harmonic Mean (MHM) estimator described in Geweke (2005). For the simple BVAR, we compute the ML using (for comparison) these same two methods and the analytical solution available under the conjugate prior and posterior.

Focussing first on the 4-variables models, the BVAR-SV features the highest ML. The BVAR-GFSV ML is smaller by 22.4 units when using the Chib method and by 6 units when using the MHM. When the additional restrictions (10) are imposed on the BVAR-GFSV, yielding the BVAR-CSV model, the ML decreases by 51.1 and 80.6 units, depending on the ML measure used. By completely removing time variation in the volatilities (BVAR), the ML further drops by 151.1 or 78.6 points, depending on the ML measure. In the 14 variable case, the BVAR-CSV features a

much higher (more than 224 log points) ML than the homoskedastic BVAR. Overall, this evidence indicates that, while the BVAR-SV model offers the best data fit, the BVAR-GFSV is not too far off from it. While the additional restrictions required for the BVAR-CSV specification reduce model fit, this model still offers huge gains in model fit compared to a conventional BVAR with constant volatility. It also offers the opportunity to include time-varying volatility in the larger VAR models that some research has found to be preferable to smaller models.

4.5 Real-time forecast results

In this subsection we assess the forecasting performance of our proposed BVAR-GFSV and BVAR-CSV models. As mentioned above, the evaluation sample is 1985Q1-2013Q4, we consider five forecast horizons, and the exercise is conducted in a real-time manner, using recursive estimation with real-time data vintages. In robustness results available upon request, we have also examined forecast accuracy for a sample of 1985Q1-2008Q2, which omits the severe portion of the Great Recession and the subsequent slow recovery. Results for the shorter sample are qualitatively very similar to those reported below for the full sample.

To assess the accuracy of first point and then density forecasts, we consider root mean square errors and average log predictive scores, respectively. To provide a rough gauge of whether differences in accuracy are significantly different, we use Diebold and Mariano (1995) *t*-statistics for equal MSE and equal average log score, applied to the forecast of each model relative to a constant volatility BVAR benchmark. We compute the tests with serial correlation-robust variances, using a rectangular kernel, $h - 1$ lags, and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997). Our use of the Diebold-Mariano test with forecasts that are, in many cases, nested is a deliberate choice. Monte Carlo evidence in Clark and McCracken (2011, 2014) indicates that, with nested models, the Diebold-Mariano test for equal MSE compared against normal critical values can be viewed as a somewhat conservative (conservative in the sense of tending to have size modestly below nominal size) test for equal accuracy in the finite sample. As most of the alternative

models can be seen as nesting the benchmark, we treat the tests as one-sided, and only reject the benchmark in favor of the null (i.e., we don't consider rejections of the alternative model in favor of the benchmark).

4.5.1 Point forecasts

For the small dataset of 4 variables, Table 4 reports the RMSE of each model relative to that of the BVAR, and the absolute RMSE for the BVAR. Hence, entries less than 1 indicate that the indicated model has a lower RMSE than the BVAR.

The results in Table 4 point to two broad findings. First, including stochastic volatility in the BVAR systematically improves the point forecasts, and in general the gains are statistically significant. For example, the BVAR-SV and BVAR-GFSV improve on the RMSE of the homoskedastic BVAR in 19 of 20 cases, and the BVAR-CSV improves on the BVAR in 18 out of 20. Most of the improvements at forecast horizons of 1, 2, and 4 quarters are significant. In the 4 cases in which the BVAR does best (all involving the federal funds rate forecasts, at 4 or 8 quarters ahead), the advantage is negligible, with the relative RMSFE falling between 1.001 and 1.007.

Second, among the models featuring variations in the volatilities, the BVAR-CSV (which is the most parsimonious specification) is the one providing the best forecasts overall. In particular the BVAR-CSV provides the best forecasts in 11 cases out of 20, while the BVAR-GFSV and the BVAR-SV provide the best forecasts in 4 and 5 cases, respectively. GDP growth and inflation provide particular examples: the BVAR-CSV model improves on the RMSE of the baseline by up to 11.9 percent for GDP growth and up to 23.7 percent for inflation, depending on the horizon.

Next, we consider forecasts based on the larger dataset of 14 variables. For a dataset of this size, the BVAR-SV and BVAR-GFSV are intractable in practice, so results can be computed only for the more parsimonious BVAR-CSV specification. Results are provided in Table 5. As with the small dataset, the evidence shows that including stochastic volatility in the BVAR systematically improves the point forecasts. The BVAR-CSV outperforms the homoskedastic BVAR in 51 cases

out of 70. This evidence is stronger at short horizons: for horizons up to one year ahead, the BVAR-CSV outperforms the homoskedastic BVAR in 35 cases out of 42. This is not so surprising, as in the long horizons the benefits implied by modeling time variation in volatilities will be attenuated since the volatilities follow stationary AR processes and therefore would converge to their unconditional means at longer horizons. At shorter horizons, the gains obtained by the BVAR-CSV can be large, going up to 10-14 percent. The cases in which the homoskedastic BVAR performs better are mostly concentrated at long horizons and for some variables (unemployment and S&P500), and the corresponding gains are small, the highest being 5.3 percent (forecasts of unemployment 1-quarter ahead).

Finally, by comparing the BVAR results in Tables 4 and 5, it is clear that the use of the larger dataset improves the point forecasts, a result in line with several contributions in the literature, such as Banbura, Giannone, and Reichlin (2010), Carriero, Clark, and Marcellino (2015), Giannone, Lenza and Primiceri (2012) and Koop (2013).

4.5.2 Density forecasts

The RMSE, while informative and commonly used for forecast comparisons, is based on the point forecasts only and therefore ignores the rest of the forecast density. Of course the introduction of drifting volatility in a VAR makes it particularly well suited for density forecasting; for a 4-variable model, Clark (2011) and Clark and Ravazzolo (2014) find that adding individual stochastic volatilities to a VAR significantly improves density forecasts. The overall accuracy of the density forecasts can be measured with log predictive density scores, motivated and described in such sources as Geweke and Amisano (2010). At each forecast origin, we compute the log predictive score using the quadratic approximation of Adolfson, Linde, and Villani. (2007).[1] Specifically, we compute the log score as:

$$s_t(y^o_{t+h}) = -0.5\ n\log(2\pi) + \log|V_{t+h|t}| + \underset{t+h}{y^o - \bar{y}_{t+h|t}}' \underset{t+h|t}{V^{-1}} \underset{t+h}{y^o - \bar{y}_{t+h|t}}, \quad (36)$$

[1] In some limited checks, we obtained qualitatively similar results with some other approaches to computing the

predictive score.

where y^o_{t+h} denotes the observed outcome, $\bar{y}_{t+h|t}$ denotes the posterior mean of the forecast distribution, and $V_{t+h|t}$ denotes the posterior variance of the forecast distribution.

For the small dataset of 4 variables, Table 6 reports differences in log scores with respect to the homoskedastic BVAR, such that entries greater than 0 indicate that the model has a better average log score (better density forecast) than the benchmark BVAR model. In these results, adding stochastic volatility to a BVAR model almost always improves the density forecasts (relative to a constant volatility BVAR), with some exceptions for 2- and 3-year ahead forecasts. The fact that the relative gains stemming from modeling time variation in volatilities are more marked at shorter forecast horizons is reasonable, as it is at short horizons that the homoskedastic specification will more likely under- or over- estimate the uncertainty around the point forecasts. Instead, the models with changing volatility can quickly update its estimate of the current volatility in the economy. This very characteristic, however, can turn into a disadvantage when forecasting at long horizons. Indeed, if (at a given forecast origin) a sharp, sudden increase in volatility happens well into in the future, the homoskedastic model has an advantage in forecasting at long horizons because by construction the model's predictive density is based on the unconditional variance over the sample, which is kept higher by periods such as the era preceding the Great Moderation that began in about 1984.

Looking at all forecast horizons, the BVAR-CSV fares relatively better than the BVAR-GFSV and the BVAR-SV, outperforming both these models in 11 out of 20 cases. At 1- and 2- quarters ahead, the BVAR-GFSV and the $BVAR - CS\ V$ systematically produce the best density forecasts, outperforming the homoskedastic BVAR and the BVAR-SV. The gains at such horizons can be substantial, often on the order of 20-30 percent and up to 58.3 percent in forecasting the federal funds rate. At longer forecast horizons, the BVAR-SV is performing better than the other drifting volatility specifications, but it is still sometimes outperformed by the homoskedastic BVAR.

We now turn our attention to results for large 14 variable dataset displayed in Table 7. As previously indicated, for such a large dataset a real-time forecasting exercise can only be implemented

for the BVAR and BVAR-CSV models. The results show that the BVAR-CSV systematically improves over the homoskedastic BVAR, with gains which are often significant. The gains are in the range of 1-10 percent for most variables, but become very high for some variables, notably the federal funds rate, for which the gain against the homoskedastic specification is 33.5 percent at 1 quarter ahead. The BVAR only rarely performs better (in just 8 cases out of 70, with the largest gain being 6 percent).

Finally, by comparing the BVAR results in Tables 6 and 7, it is clear that the use of the larger dataset generally improves the density forecasts, a result in line with Carriero, Clark, and Marcellino (2015), Giannone, Lenza and Primiceri (2012), and Koop (2013).

5 Conclusions

In this paper we first propose, in a small model setting, to model conditional volatilities as driven by a combination of a single common unobserved factor and idiosyncratic components (BVAR-GFSV model). We provide evidence that the loss in terms of marginal likelihood from assuming a common factor (rather than more generally correlated volatilities) is moderate, and that the resulting estimates of the volatilities are reasonable when compared to other models. We then consider a special case of this model in which the idiosyncratic component is eliminated and the loadings on the factor are set to 1 for all the conditional volatilities (BVAR-CSV model). Such restrictions permit a convenient Kronecker structure for the posterior variance of the VAR coefficients, which in turn allows one to estimate the model with large datasets.

Of course, the possibility of estimating a large system does not come without a cost. In particular, the restrictions discussed above do not necessarily hold in a typical dataset of macroeconomic and financial variables, especially so as the cross-sectional dimension grows. Therefore there is a trade-off between (1) using the more general BVAR-GFSV model, giving up on using the information available in a large dataset, and (2) using a model which may be more likely to be misspecified, such as the BVAR-CSV, to take advantage of large information sets.

In our empirical analysis, we show that, with a small model, while the BVAR-CSV specification does not fit the data (as measured by the marginal likelihood) as well as more general volatility specifications, it improves substantially on the fit of a BVAR with constant volatility. Moreover, the BVAR-CSV specification yields volatility estimates very similar to those obtained with the more general specifications. Similarly, with a large model, the BVAR-CSV specification fits the data much better than a conventional BVAR. We then show that the BVAR-CSV provides systematic improvements in real-time point and density forecasts compared to the standard homoskedastic BVAR when using a large dataset, while in a small dataset it has a performance at the very least comparable to both the BVAR-GFSV and the BVAR with correlated drifting volatilities similar to the specification of Primiceri (2005).

We interpret these results as evidence that our proposed BVAR-CSV specification efficiently summarizes the information in a possibly large dataset while at the same time accounts for changing volatility, which considerably improves density forecast accuracy. For these reasons this class of models should have a wide range of applicability for forecasting and possibly also for policy simulation exercises. Admittedly, however, there are likely to be limits to the size of the model for which our common factor specification yields benefits. The larger the model becomes the greater the misspecification posed by our common factor restrictions may become, such that the misspecification effects could outweigh the stochastic volatility gains. We leave to further research the exploration of the model size and variable sets for which this occurs.

6 Appendix

This appendix provides the following additional information: the prior and posterior for the BVAR-GFSV model, priors for the BVAR-SV and BVAR models, and further detail on the computational gains associated with the BVAR-CSV specification.

6.1 BVAR-GFSV priors and posteriors

Drawing on the model notation given in section 2.1, let a_j, $j = 2, \ldots, n$ denote row j of the matrix A, and collect the volatility process parameters in vectors $\beta = (\beta_1, \ldots, \beta_n)$ and $\gamma = (\gamma_{1,0}, \ldots, \gamma_{n,0}, \gamma_{1,1}, \ldots, \gamma_{n,1}, \psi)$ and the matrix $\Phi = Var(e_{1,t}, \ldots, e_{n,t}, u_t)$, whose generic element is labelled φ_j. Let Λ denote the history of Λ_t from 1 to T and **f** and **h** denote the histories of the common factor and idiosyncratic components of volatility.

We specify the following (independent) priors for the parameter blocks of the model:

$$\text{vec}(\Pi)|A \sim N(\text{vec}(\underline{\mu}_\Pi), \underline{\Omega}_\Pi), \tag{37}$$

$$a_j \sim N(\underline{\mu}_{a,j}, \underline{\Omega}_{a,j}), \; j = 2, \ldots, n, \tag{38}$$

$$\beta \sim N(\underline{\mu}_\beta, \underline{\Omega}_\beta), \tag{39}$$

$$\gamma \sim N(\underline{\mu}_\gamma, \underline{\Omega}_\gamma), \tag{40}$$

$$\varphi_j \sim IG(d_\varphi \cdot \underline{\varphi}, d_\varphi), \; j = 1, \ldots, n+1. \tag{41}$$

The prior mean and variance of Π take the common Minnesota form (without cross-variable shrinkage), using the hyperparameters described for the BVAR-CSV model For the rows a_j of the matrix A, we follow Cogley in Sargent (2005) and make the prior fairly uninformative, with prior means of 0 and variances of 10 for all coefficients. For the factor loadings β_j, $j = 1, \ldots, n$, we use a prior mean of 1 and a standard deviation of 0.5. For the slope coefficients of the log factor process ψ we use a mean of 0.9 and a standard deviation of 0.3. For the coefficients of the idiosyncratic processes $\gamma_{1,0}, \ldots, \gamma_{n,0}, \gamma_{1,1}, \ldots, \gamma_{n,1}$ we use a prior mean on the slope coefficient of 0, and a prior mean on the intercept based on OLS estimates of residual variances in the training sample. The prior standard deviation is 1 for the intercept and 0.4 for the slope coefficients, with covariance of 0. For the innovations to volatility $\varphi_1, \ldots, \varphi_{n+1}$, we use a mean of 0.01 for the common factor and of 0.03 for the idiosyncratic components, with 4 degrees of freedom for each.

Under these priors, the parameters Π, Σ, β, γ, and Φ have the following closed form conditional

posterior distributions:

$$\text{vec}(\Pi)|A,\Lambda,y \sim N(\text{vec}(\bar{\mu}_\Pi),\bar{\Omega}_\Pi), \tag{42}$$

$$a_j|\Pi,\Lambda,y \sim N(\bar{\mu}_{a,j},\bar{\Omega}_{a,j}),\ j=2,\ldots,n, \tag{43}$$

$$\beta|\Pi,A,\gamma,\Phi,\mathbf{f},\mathbf{h},y \sim N(\bar{\mu}_\beta,\bar{\Omega}_\beta), \tag{44}$$

$$\gamma|\Pi,A,\beta,\Phi,\mathbf{f},\mathbf{h},y \sim N(\bar{\mu}_\gamma,\bar{\Omega}_\gamma), \tag{45}$$

$$\varphi_j|\Pi,A,\beta,\gamma,\mathbf{f},\mathbf{f},y \sim IG\left(d_\varphi \cdot \underline{\varphi} + \sum_{t=1}^{T} v_{jt}^2\, d_\varphi + T\right),\ j=1,\ldots,n+1. \tag{46}$$

Expressions for $\bar{\mu}_\beta$ and $\bar{\mu}_\gamma$ are straightforward to obtain using standard results from the linear regression model. For the VAR coefficients, the posterior mean and variance of the conditional posterior are given by:

$$\text{vec}(\bar{\mu}_\Pi) = \bar{\Omega}_\Pi \left(\text{vec}\left(\sum_{t=1}^{T} X_t y_t' A\Lambda_t^{-1} A'\right) + \Omega_\Pi^{-1}\text{vec}(\underline{\mu}_\Pi)\right) \tag{47}$$

$$\bar{\Omega}_\Pi = \left(\Omega_\Pi^{-1} + \sum_{t=1}^{T} (A\Lambda_t^{-1} A' \otimes X_t X_t')\right), \tag{48}$$

which as discussed in Section 2.5 do not have an overall Kronecker structure.

Draws from these conditionals will constitute Gibbs sampling steps in our MCMC algorithm. The algorithm is completed by drawing the unobserved states $\alpha_t = (\ln h_{1,t}, \ln h_{2,t}, \ldots, \ln h_{n,t}, \ln f_t)'$. In particular, following the same steps as described in Section 2.6 for the BVAR-CSV model, one can derive the following observation equations for the generic j-th element of $\tilde{v}^2$:

$$\ln(\tilde{v}_{j,t}^2 + \bar{c}) = \beta_j \ln f_t + \ln h_{j,t} + \ln \varrho_{j,t}^2. \tag{49}$$

The state-space representation is completed by adding a transition equation based on equations (7) - (8). In the equations above $\ln \varrho_{j,t}^2$ is not Gaussian, but $\varrho_{j,t}$ is a Gaussian process with unit variance; therefore we can use the mixture of normals approximation of KSC to estimate volatility with a Gibbs sampler, first drawing the states of the mixture and then drawing volatility conditional on

the states.

6.2 BVAR-SV and BVAR priors

For the BVAR-SV, the prior is specified similarly to the other models. For the VAR coefficients, the prior has the Minnesota-type structure described above. For A, the prior is mostly uninformative, as described above for the BVAR-GFSV specification. For the AR processes of log volatility, the normal prior has a mean of 0.9 for the slope and intercept set to be consistent with the log of least-squares estimates of residual variances in a 40 observation training sample preceding the estimation sample. The corresponding standard deviations are set to 0.2 and 0.5 (with zero covariance). For the variance-covariance matrix of innovations to log volatility, the inverse Wishart prior has a prior mean of $0.01 \times I_n$ and $n + 2$ degrees of freedom.

For the BVAR, the prior for the VAR coefficients has the Minnesota-type structure described above. The inverse Wishart prior for the error variance-covariance matrix has a mean equal to a diagonal matrix containing residual variances from AR models fit to the data sample and $n + 2$ degrees of freedom.

6.3 Computational gains from CSV restrictions

We conclude the appendix by illustrating in more detail how the computational gains of the CSV specification arise. An intuitive way to obtain a draw of vec(Π) from (22) when the prior variance does not necessarily feature a Kronecker structure (as is the case for the BVAR-GFSV and BVAR-SV), is to generate a random vector g and compute:

$$\text{vec}(\Pi)^{draw} = \text{vec}(\bar{\mu}_\Pi) + chol[\bar{\Omega}_\Pi] \times g \tag{50}$$

where g is an $nk \times 1$ standard Gaussian vector process. The inversion of the nk-dimensional square matrix $\bar{\Omega}_\Pi^{-1}$ and the computation of its Cholesky factor $chol[\bar{\Omega}_\Pi]$ both require n^3k^3 elementary operations. Then the multiplication $chol[\bar{\Omega}_\Pi] \times g$ requires further n^3k^3 operations, for a total of $3n^3k^3$ elementary operations needed to compute a single draw of vec(Π).

If the model coefficients have a posterior variance with a Kronecker structure such as the one in (27), it is possible to dramatically reduce the number of computations. In this case, drawing a $k \times n$ matrix of standard normal random variables G, a draw of Π can be obtained as follows:

$$\Pi^{draw} = \bar{\mu}_{\Pi} + chol\left[\Omega_0^{-1} + \sum_{t=1}^{T}\left(\frac{1}{f_t}X_t X_t'\right)\right]^{-1} \times G \times chol[\Sigma]' \quad . \tag{51}$$

The scheme above involves the inversion of a k-dimensional square matrix and the computation of two Cholesky terms of size k and n respectively, which require $2k^3 + n^3$ elementary operations. Then the multiplications involve $n^2k + nk^2$ further operations, for a total of $2k^3 + n^3 + n^2k + nk^2$.

Therefore, given that $k = np + 1$, the time complexity implied by (50) is of order $O(n^6p^3)$, while that implied by (51) is of order $O(n^3p^3)$, i.e. n^3 times smaller.

References

[1] Adolfson, M., Linde, J., and Villani., M., 2007. Forecasting Performance of an Open Economy DSGE Model, *Econometric Reviews* 26, 289-328.

[2] Banbura, M., Giannone, D., and Reichlin, L., 2010. Large Bayesian Vector Autoregressions, *Journal of Applied Econometrics* 25, 71-92.

[3] Carriero A., Kapetanios, G., and Marcellino, M., 2009. Forecasting Exchange Rates with a Large Bayesian VAR, *International Journal of Forecasting* 25, 400-417.

[4] Carriero, A., Kapetanios, G., and Marcellino M., 2012. Forecasting Government Bond Yields with Large Bayesian VARs. *Journal of Banking and Finance* 36, 2026-2047.

[5] Carriero A., Clark, T. and Marcellino, M., 2013. Bayesian VARs: Specification Choices and Forecast Accuracy. *Journal of Applied Econometrics*, 30, 46-73.

[6] Chib, S., 1995. Marginal Likelihood from the Gibbs Output, Journal of the American Statistical Association, 90, 1313-1321.

[7] Chib, S., Nardari, F., and Shephard, N., 2002. Markov Chain Monte Carlo Methods for Stochastic Volatility Models, *Journal of Econometrics* 108, 281-316.

[8] Chib, S., Nardari, F., and Shephard, N., 2006. Analysis of High Dimensional Multivariate Stochastic Volatility Models, *Journal of Econometrics* 134, 341-371.

[9] Clark, T., 2011. Real-Time Density Forecasts from BVARs with Stochastic Volatility, *Journal of Business and Economic Statistics* 29, 327-341.

[10] Clark, T., and Davig, T., 2011. Decomposing the Declining Volatility of Long-Term Inflation Expectations, *Journal of Economic Dynamics and Control* 35, 981-999.

[11] Clark, T., and McCracken, M., 2011. Testing for Unconditional Predictive Ability, in *Oxford Handbook of Economic Forecasting*, Michael P. Clements and David F. Hendry, eds., Oxford: Oxford University Press.

[12] Clark, T., and McCracken, M., 2014. Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy, *Journal of Econometrics*, forthcoming.

[13] Clark, T., and Ravazzolo, F., 2014. Macroeconomic Forecasting Performance Under Alternative Specifications of Time-Varying Volatility, *Journal of Applied Econometrics*, forthcoming.

[14] Cogley, T., and Sargent, T., 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the post-WWII US, *Review of Economic Dynamics* 8, 262-302.

[15] Croushore, D., 2006. Forecasting with Real-Time Macroeconomic Data, in *Handbook of Economic Forecasting*, eds. G. Elliott, C. Granger, and A. Timmermann, Amsterdam: North Holland.

[16] Croushore, D., and Stark, T., 2001. A Real-Time Data Set for Macroeconomists, *Journal of Econometrics* 105, 111-30.

[17] D'Agostino, D., Gambetti, L., and Giannone, D., 2013. Macroeconomic forecasting and structural change, *Journal of Applied Econometrics* 28, 82-101.

[18] Del Negro, M., and Otrok, C., 2008. Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles, manuscript, Federal Reserve Bank of New York

[19] Del Negro, M., and Schorfheide, F., 2004. Priors from General Equilibrium Models for VARs, *International Economic Review* 45, 643-673.

[20] Del Negro, M., and Primiceri, G., 2014. Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum, manuscript, Northwestern University.

[21] Diebold, F., and Mariano, R., 1995. Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

[22] Durbin, J. and Koopman, S.J., 2001, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, UK.

[23] Faust, J., and Wright, J., 2009. Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset, *Journal of Business and Economic Statistics* 27, 468-479.

[24] Geweke, J., and Amisano, G., 2010. Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns, *International Journal of Forecasting* 26, 216-230.

[25] Geweke, J., 2005. *Contemporary Bayesian Econometrics and Statistics*, Wiley, Hoboken, New Jersey.

[26] Geweke, J., 1992. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, in: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), *Bayesian Statistics 4*. Clarendon Press, Oxford, pp. 641-649.

[27] Giannone, D., Lenza, M. and G. Primiceri, 2012. "Prior Selection for Vector Autoregressions, *Review of Economics and Statistics*, forthcoming.

[28] Jacquier, E., Polson, N., and Rossi, P., 1995. Stochastic Volatility: Univariate and Multivariate Extensions, Working Paper 19-95, Rodney L. White Center for Financial Research.

[29] Harvey, D., Leybourne, S. and Newbold, P., 1997. Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting* 13, 281-291.

[30] Kadiyala, K., and Karlsson, S., 1997. Numerical Methods for Estimation and Inference in Bayesian VAR-Models, *Journal of Applied Econometrics* 12, 99-132.

[31] Karlsson, S., 2013. Forecasting with Bayesian Vector Autoregression, in G. Elliott and A. Timmermann (Eds), *Handbook of Economic Forecasting*, vol. 2, Elsevier, 791-897.

[32] Kim, S., Shephard, N. and S. Chib, 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* 65, 361-393.

[33] Koop, G., 2013. Forecasting with Medium and Large Bayesian VARs, *Journal of Applied Econometrics* 28, 177-203.

[34] Koop, G., and Korobilis, D., 2013. Large Time-Varying Parameter VARs. *Journal of Econometrics* 177, 185-198.

[35] Liu, P., Mumtaz, H., and Theophilopoulou, A., 2014. The Transmission of International Shocks to the UK: Estimates Based on a Time-varying Factor Augmented VAR," Journal of International Money and Finance, 46, 1-15.

[36] Mumtaz, H., Surico, P., 2012. Evolving International Inflation Dynamics: World and Country-Specific Factors. *Journal of the European Economic Association* 10, 716-734.

[37] Omori, Y., Chib, S., Shephard, N. and J. Nakajima, 2007. Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference, *Journal of Econometrics* 140, 425-449.

[38] Osiewalski, J., and Pajor, A., 2009. Bayesian Analysis for Hybrid MSF-SBEKK Models of Multivariate Volatility, *Central European Journal of Economic Modeling and Econometrics* 1, 179-202.

[39] Pajor, A., 2006. Bayesian Analysis of the Conditional Correlation Between Stock Index Returns with Multivariate Stochastic Volatility Models, *Acta Physica Polonica B* 37, 3093-3103.

[40] Primiceri, G., 2005. Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies* 72, 821-852.

[41] Romer, C., and Romer, D., 2000. Federal Reserve Information and the Behavior of Interest Rates, *American Economic Review* 90, 429-457.

[42] Sims, C., 2002. The Role of Models and Probabilities in the Monetary Policy Process, *Brookings Papers on Economic Activity* 2, 1-40.

[43] Sims, C., 1993. A Nine-Variable Probabilistic Macroeconomic Forecasting Model, in *Business Cycles, Indicators and Forecasting*, James H. Stock and Mark W. Watson, editors, University of Chicago Press, 179-212.

[44] Sims, C., and Zha, T., 1998. Bayesian Methods for Dynamic Multivariate Models, *International Economic Review* 39, 949-68.

Table 1. Mixing and Convergence Statistics for BVAR-CSV Specifications

parameter block	# of parameters	median IF	mean IF	min IF	max IF	CD rejection rate (%)
4 variables						
Π	68	1.08	1.29	0.70	5.74	4.4
Σ	10	11.77	11.02	1.55	18.67	10.0
φ	1	6.62	6.62	6.62	6.62	0.0
ψ	1	4.71	4.71	4.71	4.71	0.0
Λ	196	5.42	5.46	3.27	7.55	0.0
14 variables						
Π	798	0.95	0.96	0.42	1.84	7.6
Σ	105	2.65	3.61	0.60	10.53	1.0
φ	1	5.06	5.06	5.06	5.06	0.0
ψ	1	2.99	2.99	2.99	2.99	0.0
Λ	196	3.14	3.19	1.57	5.44	2.6

Notes:
1. All results are based on a sample of 10,000 retained draws, obtained by sampling a total of 55,000 draws, discarding the first 5,000, and retaining every 5th draw of the post-burn sample.
2. For each individual parameter, the inefficiency factor is estimated as $1 + 2\sum_{k=1}^{\infty} \rho_k$, where ρ_k is the k-th order autocorrelation of the chain of retained draws. The estimates use the Newey-West kernel and a bandwidth of 4 percent of the sample of retained draws.
3. The last column of the table provides the percentage of parameters within each block for which Geweke's (1992) convergence diagnostic test rejects equal means, using a significance level of 10%.

Table 2. CPU time requirements for different models

model	CPU time (minutes) 10,000 draws	actual number of draws used or needed in estimation
BVAR, 4 variables, 4 lags	0.24	10,000
BVAR, 14 variables, 4 lags	32.97	10,000
BVAR-SV, 4 variables, 4 lags	4.73	105,000
BVAR-SV, 14 variables, 4 lags	505.64	105,000
BVAR-GFSV, 4 variables, 4 lags	6.02	205,000
BVAR-GFSV, 14 variables, 4 lags	517.30	205,000
BVAR-CSV, 4 variables, 4 lags	2.30	55,000
BVAR-CSV, 14 variables, 4 lags	40.17	55,000

Note: The reported CPU run times are medians across 10 different sets of model estimates (different MCMC chains), generated using version 8.3 of RATS on a 3.2 GHz Intel CPU.

Table 3. Marginal likelihoods, 1965:Q1-2013:Q4 model estimates

model	analytical solution	Chib	modified harmonic mean
4 variables			
BVAR	-1097.5	-1097.5	-1097.6
BVAR-SV	NA	-872.9	-932.4
BVAR-GFSV	NA	-895.3	-938.4
BVAR-CSV	NA	-946.4	-1019.0
14 variables			
BVAR	-5644.1	-5644.1	NA
BVAR-CSV	NA	-5419.9	NA

Table 4. Forecast RMSEs, 4-variable BVARs, 1985-2013
(RMSEs for BVAR benchmark, RMSE ratios in all others)

	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
BVAR					
GDP growth	2.622	2.872	2.883	2.545	2.312
Unemployment	0.241	0.484	0.995	1.679	1.873
GDP inflation	1.091	1.168	1.384	1.991	2.413
Fed funds rate	0.485	0.843	1.451	2.483	3.105
BVAR-SV					
GDP growth	0.930 ***	0.902 ***	0.901 **	0.924	0.981
Unemployment	0.990	0.971	0.951 *	0.958	0.979
GDP inflation	0.954 ***	0.937 ***	0.891 ***	0.854 ***	0.821 ***
Fed funds rate	0.889 ***	0.937 **	0.977	1.001	0.988
BVAR-GFSV					
GDP growth	0.934 ***	0.906 ***	0.908 **	0.937	0.992
Unemployment	0.985	0.967 *	0.951 *	0.966	0.996
GDP inflation	0.950 ***	0.936 ***	0.889 ***	0.844 ***	0.809 ***
Fed funds rate	0.887 ***	0.935 **	0.980	1.007	0.999
BVAR-CSV					
GDP growth	0.898 ***	0.881 ***	0.905 **	0.956	1.013
Unemployment	0.997	0.967	0.941 *	0.955	1.001
GDP inflation	0.944 ***	0.932 ***	0.876 ***	0.815 ***	0.763 ***
Fed funds rate	0.953 *	0.967	1.002	1.004	0.982

For the forecasts from the BVARs with common stochastic volatility, entries less than 1 indicate the model has a lower RMSE than the benchmark. To provide a rough gauge of whether the RMSE ratios are significantly different from 1, we use the Diebold-Mariano *t*-statistic for equal MSE. Differences in accuracy that are statistically different from zero are denoted by one, two, or three asterisks, corresponding to significance levels of 10%, 5%, and 1%, respectively. The underlying *p*-values are based on *t*-statistics computed with a serial correlation-robust variance, using a rectangular kernel, $h - 1$ lags, and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 5. Forecast RMSEs, 14-variable BVARs, 1985-2013

(RMSEs for BVAR benchmark, RMSE ratios in all others)

	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
BVAR					
GDP growth	2.327	2.563	2.618	2.597	2.437
Consumption	2.352	2.424	2.535	2.425	2.299
BFI	8.346	9.076	9.764	10.032	10.245
Res. inv.	13.958	16.382	17.019	15.370	15.016
Ind. prod.	3.977	5.121	5.519	5.456	5.276
Cap. util.	1.020	1.874	3.250	5.059	5.870
Employment	0.952	1.468	1.983	2.240	2.198
Hours	1.736	2.391	2.884	2.898	2.814
Unemployment	0.205	0.403	0.814	1.490	1.817
GDP inflation	1.081	1.139	1.420	1.870	2.144
PCE inflation	1.633	1.796	1.884	2.007	2.251
Fed funds rate	0.662	1.104	1.628	2.238	2.621
Term spread	0.533	0.758	0.967	1.139	1.257
S&P 500	25.872	26.537	27.495	27.554	27.353
BVAR-CSV					
GDP growth	0.962 **	0.969	1.022	0.990	0.981 *
Consumption	0.985	0.988	1.012	1.035	1.014
BFI	0.982	0.986	0.993	1.010	1.004
Res. inv.	0.976	0.942 *	0.922 **	0.930 **	0.985
Ind. prod.	0.970 **	0.970 **	0.994	0.980	0.972 *
Cap. util.	0.960 ***	0.955 ***	0.974 **	0.969	0.929
Employment	0.963 *	0.944 **	0.955	0.954	0.938 *
Hours	0.986	0.959 **	0.973	0.981	0.969 *
Unemployment	1.053	1.022	1.003	1.006	1.014
GDP inflation	0.974	0.952 ***	0.921 ***	0.917 **	0.922
PCE inflation	0.978 *	0.948 ***	0.936 ***	0.930 *	0.937
Fed funds rate	0.861 ***	0.879 **	0.945 **	1.010	1.044
Term spread	0.926 ***	0.938 **	0.996	1.028	1.071
S&P 500	0.999	1.012	1.002	1.003	1.009

Note: See the notes to Table 4.

Table 6. Average log predictive scores, 4-variable BVARs, 1985-2013
(avg. score for benchmark BVAR, differences in scores in all others)

	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
BVAR					
all variables	-4.911	-6.181	-7.872	-9.362	-9.638
GDP growth	-2.423	-2.501	-2.526	-2.487	-2.453
Unemployment	-0.035	-0.733	-1.701	-2.679	-2.627
GDP inflation	-1.530	-1.631	-1.811	-2.127	-2.309
Fed funds rate	-1.136	-1.533	-1.951	-2.382	-2.586
BVAR-SV					
all variables	0.967 ***	0.831 ***	0.673 ***	0.155	-0.157
GDP growth	0.159 ***	0.150 ***	0.127 ***	0.089 **	0.051 ***
Unemployment	0.215 ***	0.157 **	0.022	-0.180	-0.453
GDP inflation	0.061 ***	0.089 ***	0.129 ***	0.142 ***	0.187 ***
Fed funds rate	0.547 ***	0.349 ***	0.114	-0.027	-0.060
BVAR-GFSV					
all variables	1.006 ***	0.855 ***	0.581 **	-0.210	-0.757
GDP growth	0.179 ***	0.176 ***	0.145 ***	0.083	0.025
Unemployment	0.223 ***	0.143 *	-0.064	-0.435	-0.921
GDP inflation	0.061 **	0.095 ***	0.142 ***	0.133 ***	0.175 ***
Fed funds rate	0.583 ***	0.366 ***	0.090	-0.132	-0.236
BVAR-CSV					
all variables	0.930 ***	0.862 ***	0.627 ***	0.045	-0.395
GDP growth	0.203 ***	0.184 ***	0.137 ***	0.092 **	0.007
Unemployment	0.231 ***	0.172 *	-0.015	-0.233	-0.706
GDP inflation	0.067 **	0.093 **	0.150 ***	0.156 ***	0.228 ***
Fed funds rate	0.465 ***	0.332 ***	0.089	-0.136	-0.237

For the forecasts from the BVARs with common stochastic volatility, entries greater than 0 indicate the model has a better average log score (better density forecast) than the benchmark model. To provide a rough gauge of the statistical significance of differences in average log scores, we use a Diebold-Mariano *t*-test of equal average log scores. Differences in average scores that are statistically different from zero are denoted by one, two, or three asterisks, corresponding to significance levels of 10%, 5%, and 1%, respectively. The underlying *p*-values are based on *t*-statistics computed with a serial correlation-robust variance, using a rectangular kernel, $h - 1$ lags, and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 7. Average log predictive scores, 14-variable BVARs, 1985-2013
(avg. score for benchmark BVAR, differences in scores in all others)

	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
BVAR					
All variables	-27.356	-29.856	-32.620	-35.172	-36.295
GDP growth	-2.351	-2.446	-2.508	-2.550	-2.543
Consumption	-2.274	-2.324	-2.384	-2.385	-2.365
BFI	-3.507	-3.644	-3.746	-3.763	-3.789
Res. inv.	-4.066	-4.228	-4.319	-4.330	-4.329
Ind. prod.	-2.859	-3.092	-3.199	-3.233	-3.228
Cap. util.	-1.480	-2.078	-2.693	-3.179	-3.290
Employment	-1.421	-1.815	-2.136	-2.254	-2.259
Hours	-2.092	-2.342	-2.526	-2.571	-2.577
Unemployment	0.117	-0.519	-1.320	-2.161	-2.333
GDP inflation	-1.508	-1.602	-1.829	-2.096	-2.233
PCE inflation	-1.941	-2.021	-2.073	-2.173	-2.300
Fed funds rate	-1.186	-1.621	-1.999	-2.329	-2.473
Term spread	-0.796	-1.142	-1.403	-1.582	-1.668
S&P 500	-4.680	-4.725	-4.758	-4.758	-4.744
BVAR-CSV					
All variables	0.868 ***	0.893 ***	0.572 ***	0.183	0.099
GDP growth	0.071 ***	0.074 ***	0.042 *	0.068 ***	0.053 ***
Consumption	0.039 **	0.028 *	0.023	0.012	0.014
BFI	0.068 *	0.068	0.008	0.018	0.004
Res. inv.	0.135 ***	0.128 ***	0.123 ***	0.094 ***	0.052
Ind. prod.	0.077 ***	0.084 **	0.034	0.060 ***	0.043 **
Cap. util.	0.094 ***	0.089 **	0.067 *	0.073	0.066
Employment	0.086 ***	0.099 ***	0.079 **	0.071 **	0.059 **
Hours	-0.051	0.012	0.022	0.024	0.005
Unemployment	0.142 ***	0.112 **	0.067	0.065	-0.027
GDP inflation	0.064 ***	0.070 ***	0.084 ***	0.093 ***	0.101 ***
PCE inflation	0.047	0.058 *	0.046 **	0.098 ***	0.081 **
Fed funds rate	0.335 ***	0.287 ***	0.156 ***	0.045	-0.005
Term spread	0.140 ***	0.086 ***	0.009	-0.062	-0.092
S&P 500	0.059 *	0.008	-0.001	-0.008	-0.011

Note: See the notes to Table 6.

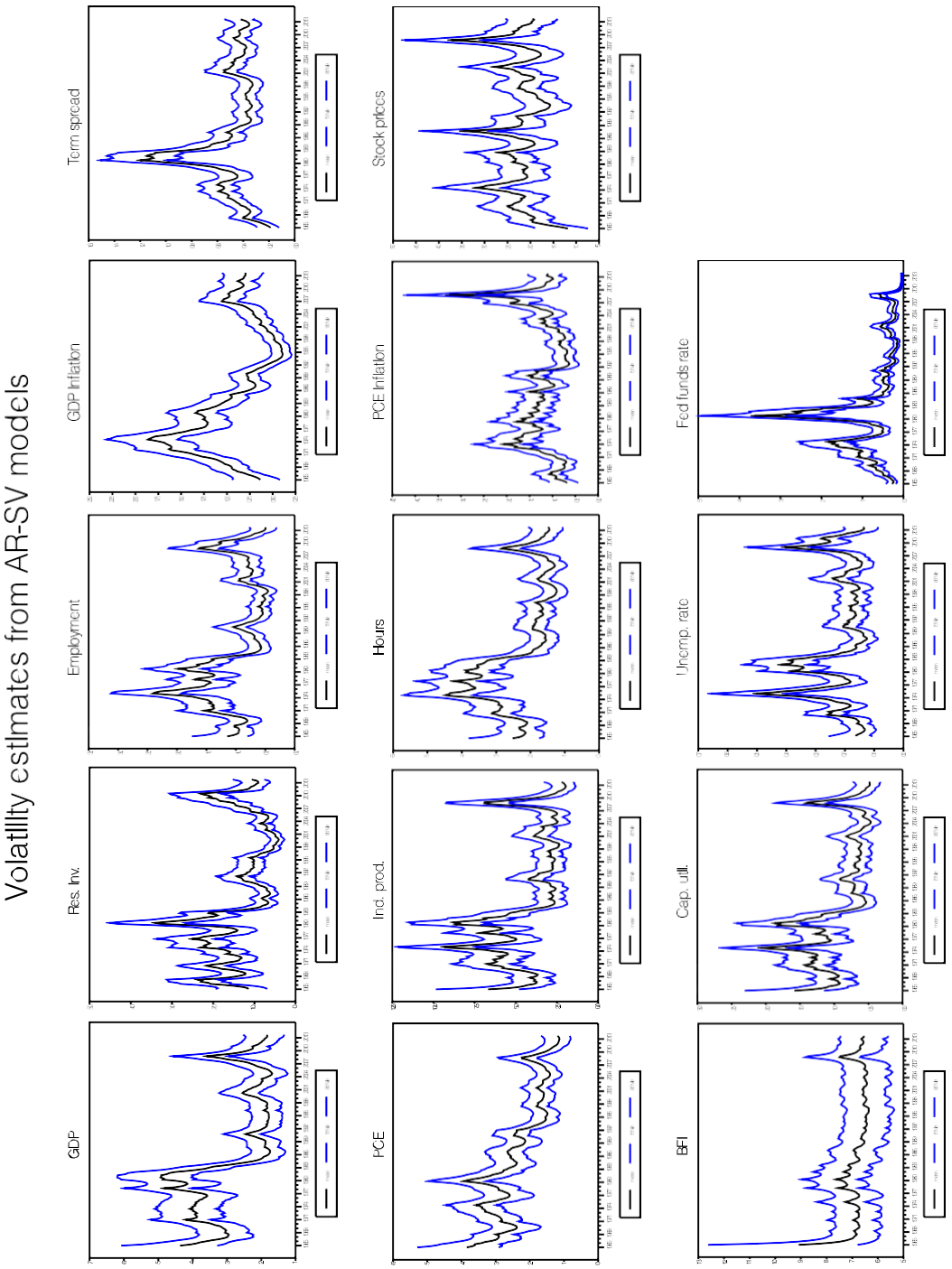

Figure 1: Volatility estimates (defined as standard deviations) from AR-SV models, final vintage data

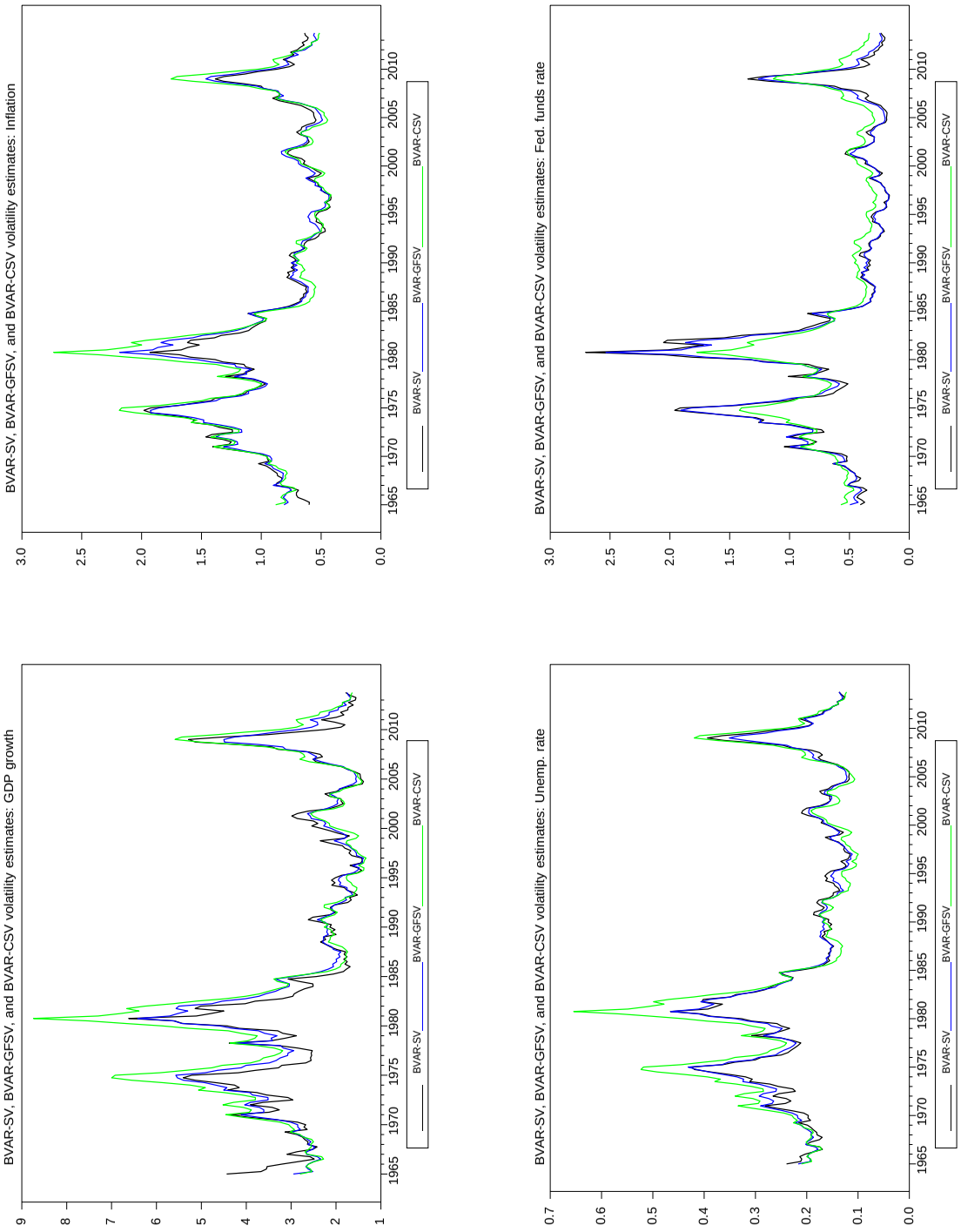

Figure 2: Volatility (defined as standard deviations) estimates from 4-variable models, final vintage data

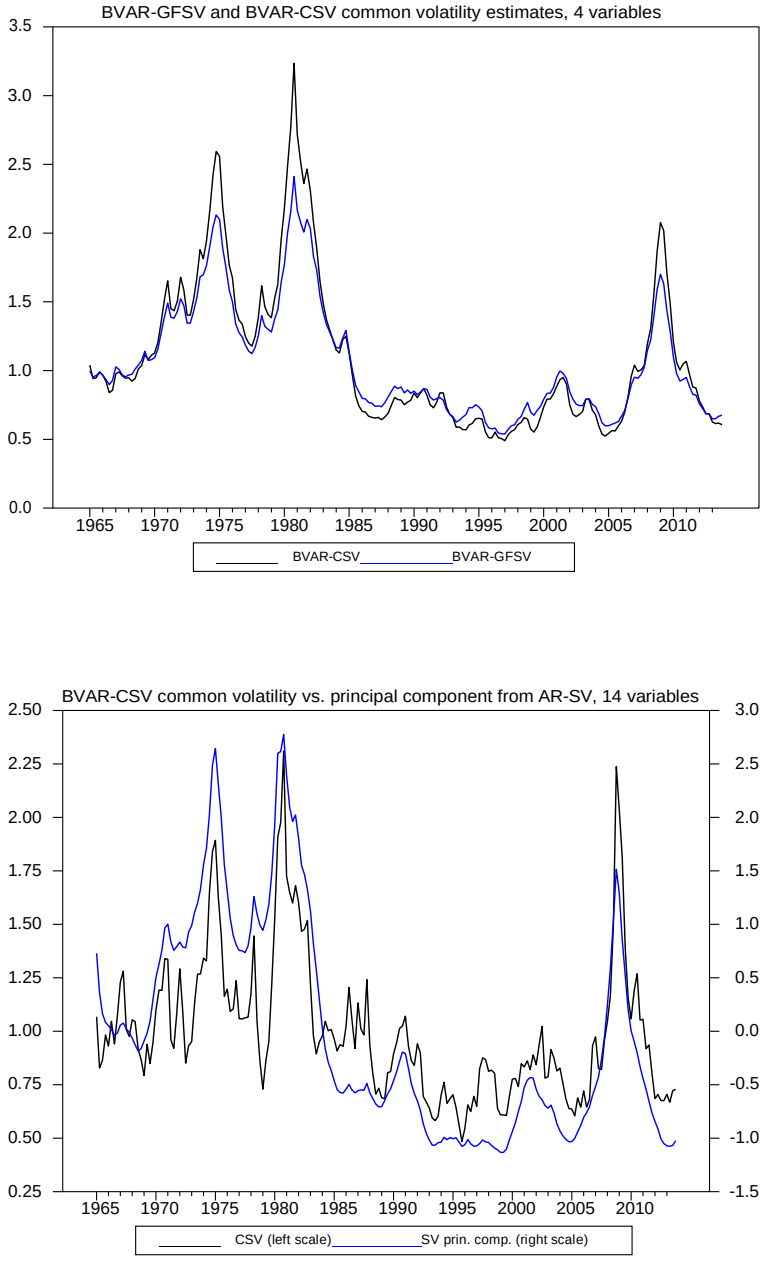

Figure 3: Top panel: Common volatility (defined as standard deviations) estimates from BVAR-GFSV and BVAR-CSV models, 4 variables, final vintage data. Bottom panel: Principal component of AR-SV estimates of volatility versus BVAR-CSV common volatility, 14 variables, final vintage data