Bayesian Compressed Vector Autoregressions*

Gary Koop
University of Strathclyde[†]

Dimitris Korobilis
University of Glasgow[‡]

Davide Pettenuzzo
Brandeis University[§]

March 23, 2016

Abstract

Macroeconomists are increasingly working with large Vector Autoregressions (VARs) where the number of parameters vastly exceeds the number of observations. Existing approaches either involve prior shrinkage or the use of factor methods. In this paper, we develop an alternative based on ideas from the compressed regression literature. It involves randomly compressing the explanatory variables prior to analysis. A huge dimensional problem is thus turned into a much smaller, more computationally tractable one. Bayesian model averaging can be done over various compressions, attaching greater weight to compressions which forecast well. In a macroeconomic application involving up to 129 variables, we find compressed VAR methods to forecast better than either factor methods or large VAR methods involving prior shrinkage.

Keywords: multivariate time series, random projection, forecasting

JEL Classifications: C11, C32, C53

*We thank Daniel Peña, Simon Price, Rob Taylor, Ruey Tsay, Herman van Dijk, Mike West, Kamil Yilmaz; conference participants in the 2016 NBER-NSF Time Series Conference and the 2016 European Seminar on Bayesian Econometrics; and seminar participants in the Bank of England, ECARES, University of Essex, University of Konstanz, University of St Andrews, for helpful comments.

†University of Strathclyde, Department of Economics, Room 4.06, Sir William Duncan Building, 130 Rottenrow, Glasgow G4 0GE, United Kingdom. gary.koop@strath.ac.uk

‡University of Glasgow, Adam Smith Business School, Gilbert Scott building, Glasgow, G12 8QQ, United Kingdom. Dimitris.Korobilis@glasgow.ac.uk

§Brandeis University, Sachar International Center, 415 South St, Waltham, MA. dpettenu@brandeis.edu

1 Introduction

Vector autoregressions (VARs) have been an important tool in macroeconomics since the seminal work of Sims (1980). Recently, many researchers in macroeconomics and finance have been using large VARs involving dozens or hundreds of dependent variables (see, among many others, Banbura, Giannone and Reichlin, 2010, Carriero, Kapetanios and Marcellino, 2009, Koop, 2013, Koop and Korobilis, 2013, Korobilis, 2013, Giannone, Lenza, Momferatou and Onorante, 2014 and Gefang, 2014). Such models often have many more parameters than observations, over-fit the data in-sample, and, as a consequence, forecast poorly out-of-sample. Researchers working in the literature typically use prior shrinkage on the parameters to overcome such over-parametrization concerns. The Minnesota prior is particularly popular, but other approaches such as the LASSO (least absolute shrinkage and selection operator, see Park and Casella, 2008 and Gefang, 2014) and SSVS (stochastic search variable selection, see George, Sun and Ni, 2008) have also been used. Most flexible Bayesian priors that result in shrinkage of high-dimensional parameter spaces rely on computationally intensive Markov Chain Monte Carlo (MCMC) methods and their application to recursive forecasting exercises can, as a consequence, be prohibitive or even infeasible. The only exception is a variant of the Minnesota prior that is based on the natural conjugate prior, an idea that has recently been exploited by Banbura, Giannone and Reichlin (2010) and Giannone, Lenza and Primiceri (2015), among others. While this prior allows for an analytical formula for the posterior, there is a cost in terms of flexibility in that a priori all VAR equations are treated in the same manner; see Koop and Korobilis (2010) for a further discussion of this aspect of the natural conjugate prior.

The themes of wishing to work with Big Data[1] and needing empirically-sensible shrinkage of some kind also arise in the compressed regression literature; see Donoho (2006). In this literature, shrinkage is achieved by compressing the data instead of the parameters. These methods are used in a variety of models and fields (e.g. neuroimaging, molecular epidemiology, astronomy). A crucial aspect of these methods is that the projections used to compress

[1] Big Data comes in two forms that are often called Tall and Fat. Tall data involves a huge number of observations, whereas Fat Data involves a huge number of variables. In this paper, we fall in the Fat Data part of the literature.

the data are drawn randomly in a data oblivious manner. That is, the projections do not involve the data and are thus computationally trivial. Recently, Guhaniyogi and Dunson (2015) introduced the idea of Bayesian Compressed regression, where a number of different projections are randomly generated and the explanatory variables are compressed accordingly. Next, Bayesian model averaging (BMA) methods are used to attach different weights to the projections based on the explanatory power the compressed variables have for the dependent variable.

In economics, alternative methods for compressing the data exist. The most popular of these is principal components (PC) as used, for instance, in the Factor-Augmented VAR, FAVAR, of Bernanke, Boivin and Eliasz (2005) or the dynamic factor model (DFM) of, e.g., Geweke (1977) and Stock and Watson (2002). PC methods compress the original data into a set of lower-dimensional factors which can then be exploited in a parsimonious econometric specification, for example, a univariate regression or a small VAR. The gains in computation from such an approach are large (but not as large as the data oblivious methods used in the compressed regression literature), since principal components are relatively easy to compute and under mild conditions provide consistent estimates of unobserved factors for a wide variety of models, including those with structural instabilities in coefficients (Bates, Plagborg-Møller, Stock and Watson, 2013). However, the data compression is done without reference to the dependent variable(s). PC is thus referred to as an unsupervised data compression method. In contrast, the methods used in the compressed regression literature, including the Guhaniyogi and Dunson (2015) approach, are supervised. To our knowledge, supervised compressed regression methods of this sort have not yet been used in the VAR literature.[2]

In this paper, we extend the Bayesian random compression methods of Guhaniyogi and Dunson (2015), developed for the regression model, to the VAR leading to the Bayesian Compressed VAR (BCVAR). In doing so, we introduce several novel features to our method. First, we generalize the compression schemes of Guhaniyogi and Dunson (2015) and apply them both to the VAR coefficients and the elements of the error covariance matrix. In

[2]Carriero, Kapetanios and Marcellino (2015) use a reduced rank VAR framework they refer to as a multivariate autoregressive index model that shares similarities with the BCVAR used in this paper. However, they use computationally-burdensome MCMC methods which would preclude their use in very high dimensional models.

high dimensional VARs, the error covariance matrix will likely contain a large number of unknown parameters. As a concrete example, the error covariance matrix of the largest VAR we considered in our empirical application includes more than 8,000 free parameters. Compressing the VAR coefficients while leaving these parameters unconstrained may still lead to a significant degree of over-parametrization and, arguably, to a poor forecast performance, which explains our desire to compress the covariance matrix. Second, we allow the explanatory variables in the different equations of the VAR to be compressed in potentially different ways. In macroeconomic VARs, where the first own lag in each equation is often found to have important explanatory power, forcing the same (compressed) variables to appear in each equation seems problematic. We accomplish this by developing a computationally efficient algorithm that breaks down the estimation of the high dimensional compressed VAR into the estimation of individual univariate regressions. Our algorithm has very low requirements in terms of memory allocation and, since the VAR equations are assumed to be independent, can be easily parallelized to fully exploit the power of modern high-performance computer clusters (HPCC).[3] Third, we generalize our compressed VAR methods to the case of large-dimensional VARs with equation-specific time-varying parameters and volatilities. This is achieved by extending the approach developed in Koop and Korobilis (2013) to the compressed VAR, relying on variance discounting methods to model, in a computationally efficient way, the time variation in the VAR coefficients and error covariance matrix.

We then carry out a substantial macroeconomic forecasting exercise involving VARs with up to 129 dependent variables and 13 lags. We compare the forecasting performance of seven key macroeconomic variables using the BCVAR to various popular alternatives: univariate AR models, the DFM, the FAVAR, and the Minnesota prior VAR (implemented as in Banbura, Giannone and Reichlin, 2010). Our results are encouraging for the BCVAR, showing substantial forecast improvements in many cases and comparable forecast performance in the remainder.

The rest of the paper is organized as follows. Section 2 provides a description of the theory behind random compression. Section 3 introduces the Bayesian Compressed VAR with constant parameters, and develops methods for posterior and predictive analysis, while

[3]This work made use of the High Performance Computing Cluster (HPC64) at Brandeis University.

section 4 describes the empirical application. Next, section 5 introduces heteroskedasticity and time-variation in the parameters of the BCVAR and documents that these extensions further improve the forecasting performance of our approach. Section 6 provides some concluding remarks.

2 The Theory and Practice of Random Compression

Random compression methods have been used in fields such as machine learning and image recognition as a way of projecting the information in data sets with a huge number of variables into a much lower dimensional set of variables. In this way, they are similar to PC methods, which take as inputs many variables and produce as the output orthogonal factors. With PC methods, the first factor accounts for as much of the variability in the data as possible, the second factor the second most, etc. Typically, a few factors are enough to explain most of the variability in the data and, accordingly, parsimonious models involving only a few factors can be constructed. Random compression does something similar, but is computationally simpler, and capable of dealing with a massively huge number of variables. For instance, in a regression context, Guhaniyogi and Dunson (2015) have an application involving 84,363 explanatory variables.

To fix the basic ideas of random compression, let X be a $T \times k$ data matrix involving T observations on k variables where $k \gg T$. X_t is the t^{th} row of X. Define the projection matrix, Φ, which is $m \times k$ with $m \ll k$ and $\widetilde{X}_t' = \Phi X_t'$. Then $\widetilde{X}_t$ is the t^{th} row of the compressed data matrix, $\widetilde{X}$. Since $\widetilde{X}$ has m columns and X has k, the former is much smaller and is much easier to work with in the context of a statistical model such as a regression or a VAR. The question is: what information is lost by compressing the data in this fashion? The answer is that, under certain conditions, the loss of information may be small. The underlying motivation for compression arises from the Johnson-Lindenstrauss lemma (see Johnson and Lindenstrauss, 1984). This states that any k point subset of Euclidean space can be embedded in $m = O\left(\log(k)/\epsilon^2\right)$ dimensions without distorting the distances between any pair of points by more than a factor of $1 \pm \epsilon$ for any $0 < \epsilon < 1$.

The random compression literature recommends treating Φ as a random matrix and drawing its elements in some fashion. A key early paper in this literature is Achlioptas

(2003), which provides theoretical justification for various ways of drawing Φ in a computationally-trivial manner. One such scheme, which we rely on in our empirical work, is to draw Φ_{ij}, the ij^{th} element of Φ, (where $i = 1,..,m$ and $j = 1,..,k$) from the following distribution:

$$\begin{array}{l} \Pr\left(\Phi_{ij} = \frac{1}{\sqrt{\varphi}}\right) = \varphi^2 \\ \Pr\left(\Phi_{ij} = 0\right) = 2\left(1-\varphi\right)\varphi \\ \Pr\left(\Phi_{ij} = -\frac{1}{\sqrt{\varphi}}\right) = \left(1-\varphi\right)^2 \end{array}, \tag{1}$$

where φ and m are unknown parameters. The theory discussed above suggests that Φ should be a random matrix whose columns have unit lengths and, hence, Gram-Schmidt orthonormalization is done on the rows of the matrix Φ.

These methods are referred to as *data oblivious*, since Φ is drawn without reference to the data. However, the statistical theory proves that even data oblivious random compression can lead to good properties. For instance, in the compressed regression model, Guhaniyogi and Dunson (2015) provide proofs of its theoretical properties asymptotically in T and k. Under some weak assumptions, the most significant relating to sparsity (e.g. on how fast m can grow relative to k as the sample size increases), Guhaniyogi and Dunson (2015) show that their Bayesian compressed regression algorithm produces a predictive density which converges to the true predictive density. The convergence rate depends on how fast m and k grow with T. With some loose restrictions on this, they obtain near parametric rates of convergence to the true predictive density. In a simulation study and empirical work, they document excellent coverage properties of predictive intervals and large computational savings relative to popular alternatives. In the large VAR, there is likely to be a high degree of sparsity since most VAR coefficients are likely to be zero, especially for more distant lag lengths. In such a case, the theoretical results of Guhaniyogi and Dunson (2015) suggest fast convergence should occur and the computational benefits will likely be large.

These desirable properties of random compression hold even for a single, data-oblivious, random draw of Φ. In practice, when working with random compressions, many random draws are taken and then averaged. For example, Guhaniyogi and Dunson (2015) rely on BMA to average across the different random projections they considered. Treating each $\Phi^{(r)}$ ($r = 1,..,R$) as defining a new model, they first calculate the marginal likelihood for each model, and then average across the various models using weights proportional to their marginal likelihoods. Note also that m and φ can be estimated as part of this BMA exercise. In fact, Guhaniyogi and Dunson (2015) recommend simulating φ from the $U[a, b]$ distribution, where a (b) is set to a number slightly above zero (below one) to ensure numerical stability. As for m, they recommend simulating it from the $U[2\log(k), \min(T, k)]$ distribution.

To see precisely how this works in a regression context, let y_t be the dependent variable and consider the regression:

$$y_t = X_t\beta + \varepsilon_t. \tag{2}$$

If $k \gg T$, then working directly with (2) is impossible with some statistical methods (e.g. maximum likelihood estimation) and computationally demanding with others (e.g. Bayesian approaches which require the use of MCMC methods). Some of the computational burden can arise simply due to the need to store in memory huge data matrices. Manipulating such data matrices even a single time can be very demanding. For instance, calculation of the Bayesian posterior mean under a natural conjugate prior requires, among other manipulations, inversion of a $k \times k$ matrix involving the data. This can be difficult if k is huge.

In order to deal with a large number of predictors, one can specify a compressed regression variant of (2)

$$y_t = (\Phi X_t)\,\beta^c + \varepsilon_t. \tag{3}$$

This model is similar to a reduced-rank regression regression (see Geweke, 1996 and Kleibergen and Van Dijk, 1998), as the k explanatory variables in the original regression model are squeezed into a small number of explanatory variables given by the vector $\widetilde{X}_t = \Phi X_t$. The crucial difference with likelihood-based approaches such as the ones proposed by Geweke (1996), Kleibergen and Van Dijk (1998), and Carriero, Kapetanios and Marcellino (2015) is that the matrix Φ is not estimated. In fact it is independent of the data and is drawn randomly using schemes such as the ones in equation (1). Once the explanatory variables have been compressed (i.e. conditional on Φ), standard Bayesian regression methods can be used for the regression of y_t on $\widetilde{X}_t$. If a natural conjugate prior is used, then analytical formulae exist for the posterior, marginal likelihood, and predictive density and computation is trivial.

It is clear that there are huge computational gains by adopting specification (3) instead of (2). In addition, the use of BMA will ensure that bad compressions (i.e. those that lead to the loss of information important for explaining y_t) are avoided or down-weighted. To provide some more intuition, note that if we were to interpret m and φ and, thus, Φ, as random parameters (instead of specification choices defining a particular compressed regression), then BMA can be interpreted as importance sampling. That is, the $U[a,b]$ and $U[2\log(k), \min(T,k)]$ distributions that Guhaniyogi and Dunson (2015) use for drawing φ and m, respectively, can be interpreted as importance functions. Importance sampling weights are proportional to the posterior for m and φ. But this is equivalent to the marginal likelihood which arises if Φ is interpreted as defining a model. Thus in this particular setting importance sampling is equivalent to BMA. In the same manner that importance sampling attaches more weight to draws from high regions of posterior probability, doing BMA with randomly compressed regressions attaches more weight to good draws of Φ which have high marginal likelihoods.

In a VAR context, doing BMA across models should only improve empirical performance since this will lead to more weight being attached to choices of Φ which are effective in explaining the dependent variables. Such supervised dimension reduction techniques contrast with unsupervised techniques such as PC. It is likely that supervised methods such as this will forecast better than unsupervised methods, a point we investigate in our empirical work.

In summary, for a given compression matrix, Φ, the huge dimensional data matrix is compressed into a much lower dimension. This compressed data matrix can then be used in a statistical model such as a regression or a VAR. The theoretical statistical literature on random compression has developed methods such as (1) for randomly drawing the compression matrix and showed them to have desirable properties under weak conditions which are likely to hold in large VARs. By averaging over different draws for Φ (which can differ both in terms of m and φ) BMA can be done. All this can be done in a computationally simple manner, working only with models of low dimension.

3 Random Compression of VARs

We start with the standard reduced form VAR model,[4]

$$Y_t = BY_{t-1} + \epsilon_t \tag{4}$$

where Y_t for $t = 1, ..., T$ is an $n \times 1$ vector containing observations on n time series variables, ϵ_t is i.i.d. $\mathcal{N}(0, \Omega)$ and B is an $n \times n$ matrix of coefficients. Note that, with $n = 100$, the uncompressed VAR will have $10,000$ coefficients in B and $5,050$ in Ω. In a VAR(13), such as the one used in this paper, the former number becomes $130,000$. It is easy to see why computation can become daunting in large VARs and why there is a need for shrinkage.

To compress the explanatory variables in the VAR, we can use the matrix Φ given in (1) but now it will be an $m \times n$ matrix where $m \ll n$, subject to the normalization $\Phi'\Phi = I$. In a similar fashion to (3), we can define the compressed VAR:

$$Y_t = B^c \left(\Phi Y_{t-1}\right) + \epsilon_t, \tag{5}$$

where B^c is $m \times n$. Thus, we can draw upon the motivations and theorems of, e.g., Guhaniyogi and Dunson (2015) to offer theoretical backing for the compressed VAR. If a natural conjugate prior is used, for a given draw of Φ the posterior, marginal likelihood, and predictive density of the compressed VAR in (5) have familiar analytical forms (see, e.g., Koop and Korobilis, 2009). These, along with a method for drawing Φ, is all that are required to forecast with the BCVAR. And, if m is small, the necessary computations of the natural conjugate BCVAR are straightforward.

We note however that the natural conjugate prior has some well-known restrictive properties in VARs.[5] In the context of the compressed VAR, working with a Φ of dimension $m \times n$ as defined in (5), with only n columns instead of n^2 would likely be much too restrictive in many empirical contexts. For instance, it would imply that to delete a variable in one equation, then that same variable would have to be deleted from all equations. In macroeconomic VARs, where the first own lag in each equation is often found to have

[4]For notational simplicity, we explain our methods using a VAR(1) with no deterministic terms. These can be added in a straightforward fashion. In our empirical work, we have monthly data and use 13 lags and an intercept.

[5]These are summarized on pages 279-280 of Koop and Koroblis (2009).

important explanatory power, such a property seems problematic. It would imply, say, that lagged inflation could either be included in every equation or none when what we might really want is for lagged inflation to be included in the inflation equation but not most of the other equations in the VAR.

An additional issue with the natural conjugate BCVAR is that it allows the error covariance matrix to be unrestricted. In high dimensional VARs, Ω contains a large number of parameters and we may want a method which allows for their compression. This issue does not arise in the regression model of Guhaniyogi and Dunson (2015) but is potentially very important in large VARs. For example, in our application the largest VAR we estimate has an error covariance matrix containing $8,385$ unknown parameters. These considerations motivate working with a re-parametrized version of the BCVAR that allows for compression of the error covariance matrix. Following common practice (see, e.g., Primiceri, 2005, Eisenstat, Chan and Strachan, 2015 and Carriero, Clark and Marcellino, 2015) we use a triangular decomposition of Ω:

$$A\Omega A' = \Sigma\Sigma, \tag{6}$$

where Σ is a diagonal matrix with diagonal elements σ_i $(i = 1, ..., n)$, and A is a lower triangular matrix with ones on the main diagonal. Next, we rewrite $A = I_n + \widetilde{A}$, where I_n is the $(n \times n)$ identity matrix and $\widetilde{A}$ is a lower triangular matrix with zeros on the main diagonal. Using this notation, we can rewrite the reduced-form VAR in (4) as follows

$$Y_t = BY_{t-1} + A^{-1}\Sigma E_t \tag{7}$$

where $E_t \sim N(0, I_n)$. Further rearranging, we have

$$\begin{aligned} Y_t &= \Gamma Y_{t-1} + \widetilde{A}(-Y_t) + \Sigma E_t \\ &= \Theta Z_t + \Sigma E_t \end{aligned} \tag{8}$$

where $Z_t = \left[Y_{t-1}', -Y_t'\right]'$, $\Gamma = AB$ and $\Theta = \left[\Gamma, \widetilde{A}\right]$. Because of the lower triangular structure of $\widetilde{A}$, the first equation of the VAR above includes only Y_{t-1} as explanatory variables, the second equation includes $\left(Y_{t-1}', -Y_{1,t}\right)'$, the third equation includes $\left(Y_{t-1}', -Y_{1,t}, -Y_{2,t}\right)'$, and so on (here $Y_{i,t}$ denotes the i-th element of the vector Y_t). Note that this lower triangular structure, along with the diagonality of Σ, means that equation-by-equation estimation of the

VAR can be done, a fact we exploit in our algorithm. Furthermore, since the elements of $\widetilde{A}$ control the error covariances, by compressing the model in (8) we can compress the error covariances as well as the reduced form VAR coefficients.

Given that in the triangular specification of the VAR each equation has a different number of explanatory variables, a natural way of applying compression in (8) is through the following specification:

$$Y_{i,t} = \Theta_i^c \left(\Phi_i Z_t^i\right) + \sigma_i E_{i,t} \quad i = 1, ..., n \tag{9}$$

where now Z_t^i denotes the subset of the vector Z_t which applies to the i-th equation of the VAR: $Z_t^1 = (Y_{t-1})$, $Z_t^2 = \left(Y_{t-1}', -Y_{1,t}\right)'$, $Z_t^3 = \left(Y_{t-1}', -Y_{1,t}, -Y_{2,t}\right)'$, and so on. Similarly, Φ_i is a matrix with m rows and column dimension that conforms with Z_t^i. Following (9), we now have n compression matrices (each of potentially different dimension and with different randomly drawn elements), and as a result the explanatory variables in the equations of the original VAR can be compressed in different ways. Note also that an alternative way to estimate a compressed VAR version of model (8) would be to write the model in its SUR form; see Koop and Korobilis (2009). Doing so implies that the data matrix Z_t would have to be expanded by taking its Kronecker product with I_n. For large n such an approach would require multiple times the memory than a modern personal computer has available. Even if using sparse matrix calculations, having to define the non-zero elements of the matrices in the SUR form of a large VAR will result in very slow computations. On the other hand, the equation-by-equation estimation we propose in (9) is simpler and can be easily parallelizable, since the VAR equations are assumed to be independent.

For a given set of posterior draws of Θ_i^c and σ_i ($i = 1, .., n$), estimation and prediction can be done in a computationally-fast fashion using a variety of methods since each model will be of low dimension and, for the reasons discussed previously, all these can be done one equation at a time. In the empirical work in this paper, we use standard Bayesian methods suggested in Zellner (1971) for the seemingly unrelated regressions model. In particular, for each equation we use the prior:

$$\begin{aligned} \Theta_i^c | \sigma_i^2 &\sim N\left(\underline{\Theta}_i^c, \sigma_i^2 \underline{V}_i\right) \\ \sigma_i^{-2} &\sim G\left(\underline{s}_i^{-2}, \underline{\nu}_i\right), \end{aligned} \tag{10}$$

where $G\left(\underline{s}_i^{-2}, \underline{\nu}_i\right)$ denotes the Gamma distribution with mean $\underline{s}_i^{-2}$ and degrees of freedom $\underline{\nu}_i$. In our empirical work, we set set $\underline{\Theta}_i^c = 0$, $\underline{V}_i = 0.5 \times I$ and, for σ_i^{-2} use the non-informative version of the prior (i.e. $\underline{\nu}_i = 0$). We then use familiar Bayesian results for the Normal linear regression model (e.g. Koop, 2003, page 37) to obtain analytical posteriors for both Θ_i^c and σ_i. The one-step ahead predictive density is also available analytically. However, h-step ahead predictive densities for $h > 1$ are not available analytically.[6] To compute them, we proceed by first converting the estimated compressed triangular VAR in equation (9) back into the triangular VAR of equation (8), noting that

$$\Theta = \left[(\Theta_1^c \Phi_1, \mathbf{0}_n)', (\Theta_2^c \Phi_2, \mathbf{0}_{n-1})', ..., (\Theta_{n-1}^c \Phi_{n-1}, 0)', (\Theta_n^c \Phi_n)'\right]' \tag{11}$$

where $\mathbf{0}_n$ is an $(1 \times n)$ vector of zeros, $\mathbf{0}_{n-1}$ is an $(1 \times n-1)$ vector of zeros, and so on. Subsequently, we go from the triangular VAR in equation (8) to the original reduced-form VAR in equation (4) by noting that $B = A^{-1}\Gamma$, where Γ can be recovered from the first $n \times n$ block of Θ in (11), and A is constructed from $\widetilde{A}$ using the remaining elements of Θ (see equation (8)). Finally, the covariance matrix of the reduced form VAR is simply given by equation (6), where both A and Σ are known. After these transformations are implemented, standard results for Bayesian VARs can be used to obtain multi-step-ahead density forecasts.

So far we have discussed specification and estimation of the compressed VAR conditional on a single compression Φ (or Φ_i, $i = 1, .., n$). In practice, we generate R sets of such compression matrices $\Phi_i^{(r)}$ ($i = 1, .., n$ and $r = 1, .., R$), and estimate an equal number of compressed VAR models, which we denote with $M_1, ..., M_R$. Then, for each model, we use the predictive simulation methods described above to obtain the full predictive density $p\left(Y_{t+h}|M_r, \mathcal{D}^t\right)$, where $h = 1, ..., H$. For each forecast horizon h, the final BMA forecast is a mixture of the form

$$p\left(Y_{t+h}|\mathcal{D}^t\right) = \sum_{r=1}^{R} w_r p\left(Y_{t+h}|M_r, \mathcal{D}^t\right), \tag{12}$$

where $\mathcal{D}^t$ is the information set available at time t, $w_r = \exp\left(-.5\Psi_r\right) / \sum_{r=1}^{R} \exp\left(-.5\Psi_r\right)$ is model M_r weight, and $\Psi_r = BIC_r - BIC_{\min}$, with BIC_r being the value of the Bayesian Information Criterion (BIC) of model M_r and $BIC_{\min}$ the minimum value of the BIC among

[6] Point forecasts can be iterated forward in the usual fashion, but predictive simulation is required to produce h-step ahead predictive densities.

all R models. We use BIC to approximate the marginal likelihood because it can be calculated easily for high-dimensional VARs and is insensitive to the choice of the priors.

In our empirical work, the $\Phi_i^{(r)}$'s are randomly drawn using the strategy described in (1). This scheme means that for each of the R random compression matrices, we have to generate the parameter φ and decide on the number of rows m of each $\Phi_i^{(r)}$ (that is, the dimension of the projected space). Both these parameters are drawn randomly: φ is drawn from the uniform $U[0.1, 0.8]$ distribution and m is drawn from the discrete $U[1, 5\ln(k_i)]$, where k_i is the number of explanatory variables included in Z_t^i for VAR equation i.[7]

We note, to conclude this section, that papers such as Achlioptas (2003) have proposed alternative schemes to the one we adopted in (1) to randomly draw the elements of Φ_i. While some of these may be potentially more efficient and can provide a higher degree of sparsity (zeros in Φ_i), in our macroeconomic application we found that a wide range of alternative random projection schemes produced almost identical forecasts. Thus, in our empirical application we will focus exclusively on the scheme proposed by Guhaniyogi and Dunson (2015), as described in equation (1).

4 Empirical Application: Macroeconomic Forecasting with Large VARs

This section introduces the macroeconomic data considered in our application and reports the forecasting performance of the Bayesian Compressed VAR methods described in section 3, relative to a number of popular alternatives. We first consider the accuracy of point forecasts, using Mean Squared Forecast Errors (MSFEs). Next, we turn to the quality of the density forecasts, and for that rely on the average of the log predictive likelihoods (ALPL), as in Geweke and Amisano (2010).

4.1 Data

We use the FRED-MD data-base of monthly US variables from January 1960 through December 2014. The reader is referred to McCracken and Ng (2015) for a description of this macroeconomic data set, which includes a range of variables from a broad range of

[7]Due to numerical stability reasons, for φ we do not consider the full support $[0, 1]$.

categories (e.g. output, capacity, employment and unemployment, prices, wages, housing, inventories and orders, stock prices, interest rates, exchange rates and monetary aggregates). We use the 129 variables for which complete data was available, after transforming all variables using the transformation codes provided in Appendix A.[8] We present detailed forecasting results for seven variables of interest: industrial production growth (INDPRO), the unemployment rate (UNRATE), total nonfarm employment (PAYEMS), the change in the Fed funds rate (FEDFUNDS), the change in the 10 year T-bill rate (GS10), the finished good producer price inflation (PPIFGS) and consumer price inflation (CPIAUCSL). In particular, we estimate VARs of different dimensions, with these seven variables included in all of our specifications. We have a Medium VAR with 19 variables, a Large VAR with 46 variables and a Huge VAR with all 129 variables. A listing of all variables (including which appear in which VAR) is given in Appendix A. Note that most of our variables have substantial persistence in them and, accordingly, the first own lag in each equation almost always has important explanatory power. Accordingly, we do not compress the first own lag. This is included in every equation, with compression being done on the remaining variables.[9]

4.2 Alternative Methods for Large VARs

We use the Bayesian compressed VAR methods introduced in section 3 in two ways: the first one, which we label as $BCVAR_C$, compresses both the VAR coefficients and the error covariances as in (9). The second one, which we label BCVAR, is the same, except for the fact that it does not compress the error covariances.

To better assess the forecasting accuracy of these compressed VAR methods, we compare their performances against a number of popular alternatives. Reasoning that previous work with large numbers of dependent variables have typically used factor methods or large VAR methods using the Minnesota prior, we focus on these. In addition, we compare the forecasts using all of these methods to a benchmark approach which uses OLS forecasts from univariate

[8] In addition to dropping a few series with missing observations, we also remove the series non-borrowed reserves, as it became extremely volatile during the Great Recession.

[9] We also standardize our variables prior to estimation and forecasting. The forecasts of the original variables are then computed by inverting the transformation and reassigning means and variances. This standardization is computed recursively, i.e., using only the data that would have been available at each point in time to estimate the various models.

AR(1) models.

Dynamic Factor Model

The dynamic factor model (DFM) can be written as:

$$\begin{aligned} Y_t &= \lambda_0 + \lambda_1 F_t + \epsilon_t \\ F_t &= \Phi_1 F_{t-1} + ... + \Phi_p F_{t-p} + \epsilon_t^F \end{aligned} \tag{13}$$

where F_t is a $q \times 1$ vector of unobserved latent factors (with $q \ll n$) which contains information extracted from all n variables, λ_0 and λ_1 are $n \times 1$ and $n \times q$ matrices, and $\epsilon_t \sim \mathcal{N}\left(0, \Sigma^Y\right)$ where Σ^Y is a diagonal matrix. The vector of factors is assumed to follow a VAR(p) process with $\epsilon_t^F \sim \mathcal{N}\left(0, \Sigma^F\right)$, with ϵ_t independent of ϵ_s^F at all t and s. We rely on principal component methods to identify the common factors.

We select the number of factors q and the lag length p as follows: We specify the maximum number of factors and lag lengths to be $q^{\max} = \sqrt{n}$ and $p^{\max} = 12$, respectively. Next, at each point in time we use BIC to choose the optimal lag length and number of factors. We use Bayesian methods with non-informative priors to estimate and forecast with this model (note that the law of motion for the common factors in equation (13) is needed to iterate forward the forecasts when $h > 1$).

Factor-Augmented VAR

We use the Factor-Augmented VAR (FAVAR) of Bernanke, Boivin, Eliasz (2005) dividing Y_t into a set of primary variables of interest, Y_t^* (these are the same key seven variables listed above), and the remainder $\widetilde{Y}_t$, and work with the model:

$$\begin{aligned} \widetilde{Y}_t &= \Lambda F_t + \epsilon_t^Y \\ \begin{bmatrix} F_t \\ Y_t^* \end{bmatrix} &= B_1 \begin{bmatrix} F_{t-1} \\ Y_{t-1}^* \end{bmatrix} + ... + B_p \begin{bmatrix} F_{t-p} \\ Y_{t-p}^* \end{bmatrix} + \epsilon_t^*. \end{aligned} \tag{14}$$

The vector $(F_t', Y_t^{*\prime})'$ is assumed to follow a VAR(p) process with $\epsilon_t^* \sim \mathcal{N}(0, \Sigma^*)$, with ϵ_t independent of ϵ_s^* at all t and s. As with the DFM model, we rely on principal component methods to extract the common factors F_t, and select the optimal number of factors q and the lag length p using BIC. We use Bayesian methods with non-informative priors to forecast with this model.

Bayesian VAR using the Minnesota Prior

We follow closely Banbura et al (2010)'s implementation of the Minnesota prior VAR which involves a single prior shrinkage parameter, ω. However, we select ω in a different manner than Banbura et al (2010), and estimate it in a data-based fashion similar to Giannone, Lenza and Primiceri (2015). We choose a grid of values for the inverse of the shrinkage factor ω^{-1} ranging from $0.5 \times \sqrt{np}$ to $10 \times \sqrt{np}$, in increments of $0.1 \times \sqrt{np}$. At each point in time, we use BIC to choose the optimal degree of shrinkage. All remaining specification and forecasting choices are exactly the same as in Banbura et al (2010) and, hence, are not reported here. In our empirical results, we use the acronym BVAR to refer to this approach.

4.3 Measures of Predictive Accuracy

We use the first half of the sample, January 1960–June 1987, to obtain initial parameter estimates for all models, which are then used to predict outcomes from July 1987 ($h = 1$) to June 1987 ($h = 12$). The next period, we include data for July 1987 in the estimation sample, and use the resulting estimates to predict the outcomes from August 1987 to July 1988. We proceed recursively in this fashion until December 2014, thus generating a time series of forecasts for each forecast horizon h, with $h = 1, ..., 12$. Note that when $h > 1$, point forecasts are iterated and predictive simulation is used to produce the predictive densities.

Next, for each of the seven key variables listed above we summarize the precision of the h-step-ahead point forecasts for model i, relative to that from the univariate AR(1), by means of the ratio of MSFEs:

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{i,j,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{bcmk,j,\tau+h}}, \tag{15}$$

where $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, and where $e^2_{i,j,\tau+h}$ and $e^2_{bcmk,j,\tau+h}$ are the squared forecast errors of variable j at time τ and forecast horizon h associated with model i ($i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$) and the AR(1) model, respectively. The point forecasts used to compute the forecast errors are obtained by averaging over the draws from the various models' h-step-ahead predictive densities. Values of $MSFE_{ijh}$ below one suggest that model i produces more accurate point forecasts than the AR(1) benchmark for variable j and forecast horizon h.

We also assess the accuracy of the point forecasts of the various methods using the multivariate loss function of Christoffersen and Diebold (1998). Specifically, we compute the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model i and the WMSFE of the benchmark AR(1) model as follows:

$$WMSFE_{ih} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{i,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{bcmk,\tau+h}}, \tag{16}$$

where $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$ are time $\tau + h$ weighted forecast errors of model i and the benchmark model, $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the (7×1) vector of forecast errors for the key series we focus on, and W is an (7×7) matrix of weights. Following Carriero, Kapetanios and Marcellino (2011), we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast.

As for the quality of the density forecasts, we follow Geweke and Amisano (2010) and compute the average log predictive likelihood differential between model i and the AR(1) benchmark,

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h}\right), \tag{17}$$

where $LPL_{i,j,\tau+h}$ ($LPL_{bcmk,j,\tau+h}$) denotes model i's (benchmark's) log predictive score of variable j, computed at time $\tau+h$, i.e., the log of the h-step-ahead predictive density evaluated at the outcome. Positive values of $ALPL_{ijh}$ indicate that for variable j and forecast horizon h on average model i produces more accurate density forecasts than the benchmark model.

Finally, we consider the multivariate average log predictive likelihood differentials between model i and the benchmark AR(1),

$$MVALPL_{ih} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(MVLPL_{i,\tau+h} - MVLPL_{bcmk,\tau+h}\right), \tag{18}$$

where $MVLPL_{i,\tau+h}$ and $MVLPL_{bcmk,\tau+h}$ denote the multivariate log predictive likelihoods of model i and the benchmark model at time $\tau + h$, computed under the assumption of joint normality.

In order to test the statistical significance of differences in point and density forecasts, we consider pairwise tests of equal predictive accuracy (henceforth, EPA; Diebold and Mariano, 1995; West, 1996) in terms of MSFE, WMSFE, ALPL, and MVALPL. All EPA tests we conduct are based on a two sided test with the null hypothesis being the AR(1) benchmark.

We use standard normal critical values. Based on simulation evidence in Clark and McCracken (2013), when computing the variance estimator which enters the test statistic we rely on Newey and West (1987) standard errors, with truncation at lag $h-1$, and incorporate the finite sample correction due to Harvey et al. (1997). In the tables, we use ***, ** and * to denote results which are significant at the 1%, 5% and 10% levels, respectively, in favor of the model listed at the top of each column.

4.4 Forecasting Results

Following Banbura et al. (2010), we choose a relatively large value for lag length ($p = 13$) for all the methods we compare, trusting in the compression or shrinkage of the various methods to remove unnecessary lags. Tables 1 through 3 and the left side of Table 7 present evidence on the quality of our point forecasts for our seven main variables of interest relative to the AR(1) benchmark. With some exceptions we are finding that BCVARs beat the benchmark and often tend to forecast better than the other approaches. This holds, with several exceptions, for every VAR dimension, variable and forecast horizon. Table 7, which presents the WMSFEs over the seven variables of interest, provides the best overall summary of our results as they relate to point forecasts. With six forecast horizons and three VAR dimensions, this table contains 18 dimensions in which point forecasts can be compared. In 16 of these, either BCVAR or BCVAR_c is the model with the lowest MSFE. In nine of these cases, compressed VAR approaches beat the benchmark in a statistically significant manner. The FAVAR is the next best approach, although it is worth noting that in some cases (e.g. with short term forecasting with the medium VAR) it does poorly, failing to beat the AR(1) benchmark.

Thus, random compression of the VAR coefficients is leading to improvements in forecast performance. Evidence relating to compression of the error covariance is more mixed. That is, usually the BCVAR_C forecasts better than the BCVAR, but there are exceptions to this pattern.

With regards to forecast horizon, no clear pattern emerges. There is a slight tendency for compressed VAR approaches to do particularly well at shorter horizons, but there are no strong differences across horizons. In terms of the individual variables, one notable pattern in these tables is that BCVAR and BCVAR_C are (with some exceptions) forecasting particularly well for the most important macroeconomic aggregates such as prices, unemployment and industrial production. It is also interesting to note that for UNEMP all of our approaches except the DFM are doing much better than the AR(1) benchmark. It is for this series in particular that the DFM is forecasting poorly. In contrast, for the long-term interest rate (GS10), our Huge or Large VAR methods are almost never beating the benchmark. But at least in this case, where small models are forecasting well, it is reassuring to see that MSFEs obtained using random compression methods are only slightly worse than the benchmark ones. This indicates that random compression methods are finding that the GS10 equation in the Huge VAR is hugely over-parametrized, but is successfully compressing the explanatory variables so as obtain results that are nearly the same as those from parsimonious univariate models.

Figures 1 through 3 present evidence on when the forecasting gains of BCVARs relative to the other approaches are achieved. These plot the cumulative sum of weighted forecasting errors (jointly for the $N = 7$ variables of interest) for the benchmark AR(1) model minus those for a competing approach, $CSWFED_{ih} = \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(we^2_{bcmk,\tau+h} - we^2_{i,\tau+h} \right)$, for different sized VAR sizes and different forecasting horizons. Positive values for this metric imply that an approach is beating the benchmark. For short horizons, BCVAR is the only approach that consistently beats the benchmark model, throughout the whole forecast period. All other approaches accumulate more forecast errors over time compared to the simple AR(1). It is particularly interesting that during the 2007-2009 crisis all multivariate methods seem to, at least temporarily, improve over the univariate AR(1). However, towards the end of the crisis, for all methods but the BCVAR relative forecast performance deteriorates abruptly. For longer forecast horizons some of the alternative multivariate models perform fairly well (e.g., at $h = 12$, the FAVAR ends up being the best model by a short margin). Nevertheless, even at these horizons BCVAR remains consistently a reliable forecasting model.

Tables 4 through 6 and the right hand side of Table 7 shed light on the quality of our density forecasts by presenting averages of log predictive likelihoods for the VARs of different dimensions. Results are similar as for MSFEs and we will not discuss them in detail. But they do differ in their strength in two ways. First, the evidence that compressed VAR approaches can beat univariate benchmarks becomes much more strong. See in particular the right hand side of Table 7 which shows strong rejection of the hypothesis of EPA at every horizon and for every VAR dimension. Second, the evidence that compressed VARs can forecast better than BVAR or FAVAR approaches becomes weaker (although DFM continues to be the worst of our approaches). Particularly with the huge VAR, standard large Bayesian VAR methods using the Minnesota prior tend to forecast slightly better than the compressed VAR approaches.

Figures 4, 5 and 6 plot the cumulative sums of the multivariate log predictive likelihood differentials, $CSMVLPLD_{ij} = \sum_{\tau=\underline{t}}^{\bar{t}-h} (MVLPL_{i,\tau+h} - MVLPL_{bcmk,\tau+h})$, for VARs of different dimensions and across a number of forecast horizons. It is interesting to note that, in contrast to Figures 1 through 3, there is not strong evidence of a large deterioration in forecasting performance relative to the univariate benchmark. In general, our compressed VAR approaches may not be best in every case, but even when they are not they are close to the best.

Finally, it is worth stressing that this section is simply comparing the forecast performance of different plausible methods for a particular data set. However, the decision whether to use compression methods should not be based solely on this forecasting comparison. In other, larger applications, plausible alternatives to random compression such as the Minnesota prior BVAR or any VAR approach which requires the use of MCMC methods, may simply be computationally infeasible. The results presented in this section show that with the present data set, random compression works fairly well. With larger data sets, it may very well be that BCVAR is the only approach that is computationally feasible.

5 Time-variation in Parameters: The Compressed TVP-SV VAR

In macroeconomic forecasting applications, it is often empirically necessary to allow for time-variation in the VAR coefficients and/or the error covariance matrix. There is an increasing literature that shows that ignoring macroeconomic volatility and possible structural changes in coefficients of a VAR can result in bad in-sample fit and poor out-of-sample forecast performance; see for example Clark (2011). Both such extensions add greatly to the computational burden since MCMC methods are usually required. In the context of the constant coefficient VAR with conjugate prior for the VAR coefficients there is a growing literature (e.g. Carriero, Clark and Marcellino, 2015, 2016 and Chan, 2015) investigating various structures for time-varying error covariance matrices which do not lead to excessively large computational demands. However, even these can be restrictive and require the use of MCMC methods which will make them unsuitable for use in extremely large models. Allowing for time-variation in the VAR coefficients (e.g. through assuming coefficients evolve according to a random walk or a Markov switching process) will also greatly increase the burden.

In this section, we show how the compressed VAR methods can be generalized to the case of a VAR with time-varying parameters and stochastic volatilities (BCVAR_{tvp}). Our model becomes

$$Y_{i,t} = \Theta^c_{i,t}\left(\Phi_i Z^i_t\right) + \sigma_{i,t} E_{i,t}. \tag{19}$$

Notice that relative to equation (9) now all parameters including the error variances may vary over time and, thus, they have t subscripts, $t = 1, ..., T$. We also remind the reader that the variables Z^i_t contain lags of the dependent variables and the terms which relate to the error covariances as defined in (8). This TVP-SV VAR model is different from the previous literature because it allows for equation by equation estimation. Papers such as Primiceri (2005) would specify the VAR in the familiar seemingly unrelated regression form, where all n VAR equations are modeled jointly. Estimation using the latter form can become cumbersome as n increases, since the posterior for both the time-varying regression coefficients and volatilities involves many manipulations involving large data matrices. Using (19), estimation of the BCVAR_{tvp} is reduced to the estimation of n univariate time-varying parameter regressions which is computationally more efficient for large n. Additionally, the possibly large matrix Z^i_t is still compressed using Φ_i as with the BCVAR.

In general, forecasting with TVP-SV VARs is computationally demanding as it typically relies on MCMC methods. In our case, even if we use Φ_i to compress the data, a full Bayesian analysis could be computationally demanding with large n since MCMC methods are required and must be run for each of the n equations. Accordingly, we turn to approximate methods to deal with the TVP-SV aspect of our BCVAR. These are generalizations of those developed by Koop and Korobilis (2013) in the context of a time-varying parameter with time varying error covariance matrix. They use variance discounting methods to model the time-variation in the

VAR coefficients and error covariance matrix, and provide analytical formulae for updating them. Thus, in (19), once we draw Φ_i randomly, $\Theta^c_{i,t}$ and $\sigma^2_{i,t}$ can be updated using simple recursive formulae based on the Kalman filter, without relying on computationally intensive MCMC methods.

Adapting Koop and Korobilis (2013), the compressed TVP-SV VAR model involves estimating $\Theta^c_{i,t}$ and $\sigma^2_{i,t}$ by assuming that they evolve according to:

$$\Theta^c_{i,t} = \Theta^c_{i,t-1} + \sqrt{\frac{(1-\lambda_{i,t})\, var\left(\Theta^c_{i,t|t-1}\right)}{\lambda_{i,t}}}\, u_{i,t}, \tag{20}$$

$$\sigma^2_{i,t} = \kappa_{i,t}\sigma^2_{i,t-1} + (1-\kappa_{i,t})\, \widetilde{E}^2_{i,t}. \tag{21}$$

That is, $\Theta^c_{i,t}$ follows a random walk using a forgetting factor approximation to its error covariance matrix. Kalman filtering methods can be used for this equation. For $\sigma^2_{i,t}$ we have an Exponentially Weighted Moving Average filter. $\widetilde{E}^2_{i,t}$ is the time t prediction error estimated from the i-th equation of the VAR, $u_{i,t} \sim N(0,1)$, and $var\left(\Theta^c_{i,t|t-1}\right)$ is the filtered variance of $\Theta^c_{i,t}$ given information up to time $t-1$ and is produced by the Kalman filter (see Koop and Korobilis, 2013, for details). The crucial parameters in this specification are the forgetting and decay factors $\lambda_{i,t}$ and $\kappa_{i,t}$. These factors, which are typically in the range of $(0.9, 1)$, control how quickly discounting of past data occurs. For example, if $\lambda_{i,t} = 0.90$ then $\Theta^c_{i,t}$ depends very heavily on recent observations, and changes very rapidly over time. On the other hand, if $\lambda_{i,t} = 0.99$ the discounting of the past is more gradual and $\Theta^c_{i,t}$ varies more smoothly. Finally, when $\lambda_{i,t} = 1$ we go back to the constant parameter VAR. Similar arguments can be made for $\sigma^2_{i,t}$ and its decay factor $\kappa_{i,t}$.

For out-of-sample forecasting, we extend the methods of Koop and Korobilis (2013) by allowing for the decay and forgetting factors to vary over time using simple updating formulae:

$$\lambda_{i,t} = \underline{\lambda} + (1-\underline{\lambda}) \times \exp\left(-0.5 \times \frac{\widetilde{E}^2_{i,t-1}}{\widehat{\sigma}^2_{i,t-1}}\right), \tag{22}$$

$$\kappa_{i,t} = \underline{\kappa} + (1-\underline{\kappa}) \times \exp\left(-0.5 \times kurt\left(\widetilde{E}_{i,t-12:t-1}\right)\right), \tag{23}$$

where $\widehat{\sigma}^2_{i,t-1}$ is the time $t-1$ estimate of the variance and $kurt\left(\widetilde{E}_{i,t-12:t-1}\right)$ is the excess kurtosis of the VAR prediction error, evaluated over the past year (i.e. with monthly data this is based on a rolling sample of 12 observations). $\underline{\lambda}$ and $\underline{\kappa}$ put bounds on the minimum values of the forgetting and decay factors. We set $\underline{\lambda} = 0.98$ and $\underline{\kappa} = 0.94$ which, in the context of monthly data, allow for the possibility of a fairly large amount of time variation.[10]

Note that, if the prediction error is close to zero then $\lambda_{it} = 1$ which is the value consistent with the parameters in equation i being constant. In words, if the model forecast well last month, we do not change its parameters this month. However, the larger the prediction error is, the smaller λ_{it} becomes and, thus, a higher degree of parameter change is allowed for. For the decay factor $\kappa_{i,t}$, we use a similar reasoning, except in terms of the excess kurtosis of the prediction error. As is well known (e.g. from the GARCH literature), assuming that errors are Normally distributed, in times of constant volatility excess kurtosis will be equal to zero, but in times of increased volatility kurtosis is higher. Allowing for κ_{it} to depend on the excess kurtosis over the past year is a simple way of allowing $\sigma_{i,t}$ to change more rapidly in unstable times than in stable times. Using these methods, it is straightforward to allow for time-variation in our compressed VAR approach in a computationally simple manner.

Figure 7, which plots the time series of the predictive density volatilities for the Medium BCVAR$_{tvp}$ against the time series of volatilities obtained from the alternative methods described in section 4, confirms that heteroskedasticity plays a very important role in our data. While the alternative methods allow for some time variation in the volatilities (they are estimated on an expanding window of data), BCVAR$_{tvp}$ is finding a lot more variation. This is particularly true at the time of the financial crisis.

Table 8, Figure 8, and Figure 9 present results on the forecast performance of our BCVAR$_{tvp}$ approach. The story that jumps out is a strong one: adding time variation in the parameters and volatilities leads to substantial improvements in forecast performance. Conventional wisdom has it that allowing for time-variation (particularly in the error covariance matrix) is particularly important for predictive density estimation. In a time of fluctuating volatility, working with a homoskedastic model may not seriously affect point forecasts, but may lead to poor estimates of higher predictive moments. This wisdom is strongly reinforced by our results. The right panels of Table 8 show that in terms of

[10]The idea of allowing the value of the forgetting factor to depend on the most recent prediction error is used, e.g., in Park, Jun, and Kim (1991).

predictive likelihoods, the BCVAR_{tvp} performs much better than our other compressed VAR approaches, and bettere (with some exceptions) than standard large VAR and factor methods. This is particularly true when focusing on the multivariate predictive performance and short to medium forecast horizons. In addition, improvements relative to the univariate benchmark (as indicated by the stars in the table) are always strongly statistically significant. In terms of MSFEs, allowing for time variation in parameters leads to some improvements, but these improvements are not as large as those we find with predictive likelihoods. Again, the multivariate results are particularly strong, for all VAR sizes and forecast horizons. In summary, the message conveyed by Table 8 is a particularly strong one: BCVAR_{tvp} is forecasting better than any other approach considered in this paper.

Figure 8 indicates that, with some exceptions, the reported success in terms of overall point forecast accuracy of the BCVAR_{tvp} relative to the alternative methods we considered (namely, DFM, FAVAR, and BVAR) is not the result of any specific and short-lived episodes but is instead built gradually throughout the forecast evaluation period, as indicated by the increasing lines depicted in the figure. Interestingly, both at $h = 1$ and $h = 12$, the improvements in forecast performance relative to the various alternatives are particularly notable around the time of the financial crisis, but are not confined to it. Figure 9 provides a similar analysis in terms of the overall density forecast accuracy of the BCVAR_{tvp} model. The left panels of the figure show that at $h = 1$ once again the previously reported forecast success of the BCVAR_{tvp} is built steadily throughout the forecast evaluation period. On the other hand, the right panels of the figure, which focus on $h = 12$, show that while up to the beginning of the latest financial crisis the BCVAR_{tvp} is forecasting more accurately than all the alternatives, the 2007-2009 period has a strong negative impact on the density forecast performance of the BCVAR_{tvp}.

6 Conclusions

In this paper, we have drawn on ideas from the random projection literature to develop methods suitable for use with large VARs. For such methods to be suitable, they must be computationally simple, theoretically justifiable and empirically successful. We argue that the BCVAR methods developed in this paper meet all these goals. In a substantial macroeconomic application, involving VARs with up to 129 variables, we find BCVAR methods to be fast and yield results which are at least as good as or better than competing approaches. And, in contrast to the Minnesota prior BVAR, BCVAR methods can easily be scaled up to much higher dimensional models and extended to allow for time-variation in its parameters.

References

Achlioptas, D. (2003). Database-friendly random projections: Johnson-Lindenstrauss with binary coins. Journal of Computer and System Sciences, 66, 671–687.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70, 191-221.

Banbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions. Journal of Applied Econometrics, 25, 71-92.

Bates, B. J., Plagborg-Møller, M., Stock, J. H. and Watson, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability. Journal of Econometrics, 177, 289-304.

Bernanke, B., Boivin, J., Eliasz, P. (2005). Measuring monetary policy: A factor augmented vector autoregressive (FAVAR) approach. Quarterly Journal of Economics 120, 387-422.

Carriero, A., Clark, T. and Marcellino, M. (2015). Large vector autoregressions with asymmetric priors and time varying volatilities, manuscript.

Carriero, A., Clark, T. and Marcellino, M. (2016). Common drifting volatility in large Bayesian VARs. Journal of Business & Economic Statistics, forthcoming.

Carriero, A., Kapetanios, G. and Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. International Journal of Forecasting, 25, 400-417.

Carriero, A., Kapetanios, G. and Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. Journal of Applied Econometrics, 26, 735-761.

Carriero, A., Kapetanios, G. and Marcellino, M. (2015). Structural analysis with multivariate autoregressive index models. Centre for Economic Policy Research Discussion Paper DP10801.

Chan, J. (2015). Large Bayesian VARs: A flexible Kronecker error covariance structure, manuscript.

Clark, T. E. (2011). Real-Time density forecasts from Bayesian vector autoregressions With Stochastic Volatility. Journal of Business and Economic Statistics, 29(3), 327-341.

Christoffersen, P. F. and Diebold, F. X. (1998). Cointegration and long-horizon forecasting. Journal of Business and Economic Statistics, 16(4), 450-458.

Donoho, D. (2006). Compressed sensing. IEEE Transactions on Information Theory, 52(4), 1289-1306.

Eisenstat, E., Chan, J. and Strachan, R. (2015). Stochastic model specification search for time-varying parameter VARs, Econometric Reviews, forthcoming.

Gefang, D. (2014). Bayesian doubly adaptive elastic-net lasso for VAR shrinkage. International Journal of Forecasting, 30, 1-11.

George, E., Sun, D. and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. Journal of Econometrics, 142, 553-580.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. J. Aigner and A. S. Goldberger (Eds.), Latent variables in socio-economic models, Amsterdam: North-Holland.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. Journal of Econometrics, 75(1), 121-146.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. International Journal of Forecasting, 26(2), 216-230.

Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2014). Short-term inflation projections: a Bayesian vector autoregressive approach. International Journal of Forecasting, 30, 635-644.

Giannone, D., Lenza, M. and Primiceri, G. (2015). Prior selection for vector autoregressions. Review of Economics and Statistics, 97, 436-451.

Guhaniyogi, R. and Dunson, D. (2015). Bayesian compressed regression. Journal of the American Statistical Association, 110, 1500-1514.

Johnson, W. and Lindenstrauss, J. (1984). Extensions of Lipshitz mapping into Hilbert space. Contemporary Mathematics, 26, 189-206.

Kadiyala, K. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR models. Journal of Applied Econometrics, 12, 99-132.

Koop, G. (2008). Bayesian Econometrics. Chichester: John Wiley and Sons.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. Journal of Applied Econometrics, 28, 177-203.

Koop, G. and Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. Foundations and Trends in Econometrics, 3, 267-358.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. Journal of Econometrics, 177, 185-198.

Koop, G. and Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. The Econometrics Journal, 7, 550-565.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. Journal of Applied Econometrics, 28, 204-230.

McCracken, M. and Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. Federal Reserve Bank of St. Louis, working paper 2015-012A.

Park, T. and Casella, G. (2008). The Bayesian lasso. Journal of the American Statistical Association, 103, 681-686.

Primiceri. G., (2005). Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72, 821-852.

Sims, C. (1980). Macroeconomics and reality. Econometrica, 48, 1-48.

Stock, J. and Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics, 20,147-162.

Kleibergen, F. and Van Dijk, H. (1998). Bayesian simultaneous equations analysis using reduced rank structures. Econometric Theory, 14, 701-743.

Zellner, A. (1971). An introduction to Bayesian inference in econometrics. New York: John Wiley and Sons.

Tables and Figures

Table 1. Out-of-sample point forecast performance, Medium VAR

Variable	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
			$h=1$					$h=2$		
PAYEMS	1.142	1.041	0.892	**0.799*****	0.813***	0.957	0.871	**0.571*****	0.688***	0.711***
CPIAUCSL	1.131	1.102	**0.925**	0.951	0.935**	1.062	1.115	0.964	0.940	**0.925****
FEDFUNDS	2.044	2.043	2.727	1.027	**0.952**	1.327	1.691	2.437	**0.999**	1.002
INDPRO	0.892**	0.896*	**0.821****	0.830***	0.894***	0.935	0.997	**0.835***	0.914*	0.934
UNRATE	23.061	**0.758*****	0.764**	0.762***	0.798***	9.696	0.626**	**0.582****	0.626***	0.662***
PPIFGS	0.999	0.996	0.968	**0.966**	0.982	1.041	1.050	1.060	1.030	**1.000**
GS10	1.133	**0.973**	1.095	1.007	0.997	1.027	1.032	1.082	**0.995**	1.002
			$h=3$					$h=6$		
PAYEMS	0.894	0.827	**0.530*****	0.630***	0.649***	0.971	0.825	**0.687***	0.694**	0.703**
CPIAUCSL	1.071	1.083	0.993	0.969	**0.956**	1.022	0.976	0.983	**0.962**	0.967
FEDFUNDS	1.218	1.372	1.862	1.063	**1.030**	1.073	0.998	1.230	**0.976**	0.987
INDPRO	0.951	0.959	0.924	**0.917***	0.939	0.965	0.980	1.022	0.965	**0.963**
UNRATE	5.617	0.620**	**0.526****	0.563***	0.603***	2.189	0.637*	**0.512***	0.535**	0.575**
PPIFGS	1.033	1.034	1.070	1.051	**1.026**	1.035	**1.021**	1.092	1.046	1.042
GS10	1.041	1.040	1.144	1.058	**1.031**	**1.014**	1.019	1.123	1.054	1.038
			$h=9$					$h=12$		
PAYEMS	0.997	0.850	0.798	**0.779***	0.798*	1.012	**0.885**	0.901	0.887	0.893
CPIAUCSL	0.997	0.958*	0.968	0.956	**0.943**	1.006	0.972	0.985	**0.966**	0.970
FEDFUNDS	1.054	0.926	1.065	**0.904**	0.925	1.056	**0.925**	1.135	0.960	0.990
INDPRO	**0.943**	0.952	1.022	0.953	0.969	**0.947**	0.965	0.992	0.965	0.975
UNRATE	1.446	0.645	0.569*	**0.562****	0.601**	1.122	0.661*	0.640*	**0.601****	0.636**
PPIFGS	1.014	**0.994**	1.075	1.051	1.031	1.024	**1.001**	1.105	1.062	1.046
GS10	1.018	**1.006**	1.047	1.023	1.016	1.020	**1.006**	1.054	1.031	1.019

This table reports the ratio between the MSFE of model i and the MSFE of the benchmark AR(1) for the Medium size VAR, computed as

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{i,j,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{bcmk,j,\tau+h}},$$

where $e^2_{i,j,\tau+h}$ and $e^2_{bcmk,j,\tau+h}$ are the squared forecast errors of variable j at time τ and forecast horizon h generated by model i and the AR(1) model, respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$, $j \in \{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$, and $h \in \{1, 2, 3, 6, 9, 12\}$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 2. Out-of-sample point forecast performance, Large VAR

Variable	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
			$h=1$					$h=2$		
PAYEMS	1.273	0.906	**0.788****	0.888***	0.905**	0.897	0.698***	**0.521*****	0.805***	0.837***
CPIAUCSL	1.129	1.100	1.009	0.998	**0.953**	1.160	1.118	1.110	0.943	**0.909****
FEDFUNDS	2.387	1.671	2.461	**1.093**	1.103	2.133	1.415	2.594	**0.991**	1.150
INDPRO	0.870*	0.890**	**0.783*****	0.838***	0.914***	0.861	0.966	**0.772****	0.934*	0.929*
UNRATE	30.779	**0.786*****	0.824*	0.798***	0.855**	14.267	**0.661*****	0.666*	0.722***	0.758**
PPIFGS	1.042	1.013	1.045	0.987	**0.986**	1.151	1.056	1.162	1.012	**1.004**
GS10	1.001	**0.983**	1.106	1.006	0.992	1.029	**0.987**	1.149	1.052	1.044
			$h=3$					$h=6$		
PAYEMS	0.814	0.710***	**0.489*****	0.757***	0.748***	0.823	0.832**	**0.647***	0.773**	0.762***
CPIAUCSL	1.134	1.067	1.154	0.948	**0.913****	1.055	0.970	0.999	0.910**	**0.902****
FEDFUNDS	1.878	**1.026**	2.241	1.034	1.108	1.433	**0.944**	1.224	1.098	1.088
INDPRO	0.925	0.943	**0.862**	0.957*	0.952	**0.942**	0.962	0.980	0.959	0.984
UNRATE	8.547	0.631***	**0.615***	0.677***	0.731**	3.442	0.663***	**0.617**	0.671***	0.712**
PPIFGS	1.163	1.018	1.177	1.021	**1.013**	1.129	1.014	1.095	1.012	**0.998**
GS10	1.048	**1.030**	1.222	1.057	1.059	1.040	**1.018**	1.115	1.043	1.029
			$h=9$					$h=12$		
PAYEMS	0.895	0.930	**0.840**	0.865	0.844**	**0.919**	0.966	0.999	0.972	0.940
CPIAUCSL	1.055	0.975	0.932	0.887**	**0.867*****	1.065	0.984	0.904	0.902**	**0.879*****
FEDFUNDS	1.250	**0.999**	1.139	1.060	1.028	1.131	**0.994**	1.259	1.092	1.041
INDPRO	0.983	**0.972**	1.018	1.004	0.999	**0.954**	0.979	1.056	1.002	1.020
UNRATE	2.129	**0.684*****	0.715	0.698**	0.739**	1.595	**0.710*****	0.831	0.728**	0.756**
PPIFGS	1.068	1.000	1.051	**0.985**	0.992	1.101	1.002	1.039	1.004	**0.972**
GS10	1.008	**1.001**	1.050	1.009	1.019	1.015	**1.001**	1.054	1.023	1.014

This table reports the ratio between the MSFE of model i and the MSFE of the benchmark AR(1) for the Large size VAR, across a number of different forecast horizons h. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ and $h \in \{1, 2, 3, 6, 9, 12\}$. See notes under Table 1 for additional details.

Table 3. Out-of-sample point forecast performance, Huge VAR

Variable	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
			$h=1$					$h=2$		
PAYEMS	0.772**	0.956	**0.702*** **	0.761***	0.779***	0.677**	0.679**	**0.447*** **	0.641***	0.661***
CPIAUCSL	0.937	0.948	**0.872***	0.947	0.943	0.996	1.003	0.939	0.890**	**0.879** **
FEDFUNDS	2.070	1.701	1.984	**0.891**	0.927	1.832	1.472	2.187	**0.929**	0.961
INDPRO	0.836**	0.830***	**0.785*** **	0.856***	0.913***	0.860	0.844*	**0.786** **	0.924**	0.939
UNRATE	9.263	**0.736*** **	0.883	0.781***	0.813***	4.428	**0.558*** **	0.646**	0.644***	0.703***
PPIFGS	0.965	0.985	**0.947**	0.985	1.004	1.063	1.051	1.071	1.029	**1.016**
GS10	1.106	**0.995**	1.100	0.998	1.019	1.064	1.057	1.150	**1.024**	1.032
			$h=3$					$h=6$		
PAYEMS	0.689*	0.616**	**0.412*** **	0.591***	0.591***	0.893	0.742*	**0.516** **	0.645**	0.662**
CPIAUCSL	0.968	0.984	0.978	**0.903**	0.923	0.940	0.910	1.019	0.916	**0.893**
FEDFUNDS	1.596	1.182	1.848	**0.961**	0.985	1.363	**0.957**	1.275	0.995	0.986
INDPRO	0.934	0.917	**0.865**	0.951	0.947	1.027	**0.952**	0.972	0.961	0.970
UNRATE	2.680	**0.495*** **	0.540**	0.590***	0.643***	1.068	0.479**	**0.435** **	0.562***	0.601***
PPIFGS	1.086	**1.032**	1.089	1.040	1.043	1.093	**1.037**	1.130	1.058	1.055
GS10	**1.051**	1.104	1.212	1.058	1.057	1.060	1.047	1.176	1.053	**1.033**
			$h=9$					$h=12$		
PAYEMS	1.002	0.863	**0.649**	0.744*	0.748*	1.067	0.949	**0.788**	0.854	0.842
CPIAUCSL	0.918	0.883**	0.994	0.876	**0.871** **	0.914	0.904***	0.992	0.880*	**0.862** **
FEDFUNDS	1.226	**0.950**	1.066	0.960	0.973	1.199	1.024	1.141	1.037	**1.007**
INDPRO	1.034	0.992	1.042	**0.973**	0.990	**0.981**	0.981	1.056	0.990	0.985
UNRATE	0.778	0.512***	**0.448** **	0.577***	0.610***	0.691	0.555***	**0.497** **	0.603***	0.640***
PPIFGS	1.052	**1.003**	1.127	1.062	1.019	1.081	**1.023**	1.159	1.069	1.037
GS10	1.026	1.013	1.070	1.025	**1.003**	1.049	1.021	1.084	1.035	**1.016**

This table reports the ratio between the MSFE of model i and the MSFE of the benchmark AR(1) for the Huge size VAR, across a number of different forecast horizons h. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ and $h \in \{1, 2, 3, 6, 9, 12\}$. See notes under Table 1 for additional details.

Table 4. Out-of-sample density forecast performance, Medium VAR

Variable	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
			$h=1$					$h=2$		
PAYEMS	0.058**	0.050**	**0.205*****	0.084***	0.081***	0.114***	0.118***	**0.360*****	0.165***	0.162***
CPIAUCSL	-0.210	-0.424	-0.612	0.031	**0.091****	-0.134	-0.437	-1.516	-0.146	**-0.008**
FEDFUNDS	-0.013	0.038***	**0.131*****	-0.015	-0.014	0.033**	0.021	**0.116****	0.001	-0.004
INDPRO	0.083***	0.122**	-0.084	**0.151***	0.010	-0.041	0.064	-0.027	**0.084****	0.043**
UNRATE	-1.658	0.146***	**0.167*****	0.115***	0.093***	-1.270	0.229***	**0.325*****	0.217***	0.213**
PPIFGS	-0.147	0.095	-0.353	**0.205**	0.055	-0.206	-0.251	-0.780	**-0.127**	-0.149
GS10	-0.038	**0.019**	0.010	-0.013	-0.009	-0.010	**-0.007**	-0.011	-0.011	-0.014
			$h=3$					$h=6$		
PAYEMS	0.116***	0.114***	**0.362*****	0.198***	0.189***	0.046	0.092*	**0.246*****	0.172***	0.174***
CPIAUCSL	**-0.023**	-0.061	-1.090	-0.161	-0.027	**0.110**	0.033	-0.760	-0.102	-0.081
FEDFUNDS	0.021	0.019*	**0.112*****	-0.008	-0.014	0.018**	0.017**	**0.117*****	-0.001	-0.005
INDPRO	-0.024	**0.085**	-0.013	0.066**	0.066**	0.086	0.147	-0.070	0.030***	**0.240**
UNRATE	-0.886	0.332*	**0.459****	0.410*	0.403*	0.147	0.781	0.637*	**0.951**	0.913
PPIFGS	-0.101	-0.084	-0.482	-0.105	**0.101**	**0.168**	0.127	-0.644	-0.055	-0.015
GS10	**-0.008**	-0.011	-0.010	-0.020	-0.016	0.014	**0.016**	0.006	-0.010	-0.008
			$h=9$					$h=12$		
PAYEMS	0.024	0.082	**0.131****	0.123***	0.122***	-0.003	0.031*	0.042	0.064***	**0.074*****
CPIAUCSL	-0.011	**0.000**	-0.794	-0.259	-0.265	-0.021	**0.033**	-0.783	-0.168	-0.039
FEDFUNDS	0.012*	0.008	**0.115*****	-0.003	-0.010	0.006	0.007	**0.116*****	-0.010	-0.015
INDPRO	-0.002	-0.049	-0.103	**0.057**	-0.017	**0.092**	-0.044	-0.053	0.007	0.043
UNRATE	0.905	1.292	0.688	**1.297**	1.102	2.139	**2.397**	1.330	2.108	1.715
PPIFGS	-0.051	**0.086**	-0.390	0.021	0.050	**0.075**	-0.021	-0.392	-0.007	0.002
GS10	0.002	0.002	**0.032***	-0.009	-0.028	-0.014	-0.015	**0.003**	-0.019	-0.028

This table reports the average log predictive likelihood (ALPL) differential between model i and the benchmark AR(1) for the Medium size VAR , computed as

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h} \right),$$

where $LPL_{i,j,\tau+h}$ and $LPL_{bcmk,j,\tau+h}$ are the log predictive likelihoods of variable j at time τ and forecast horizon h generated by model i and the AR(1) model, respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$, $j \in \{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$, and $h \in \{1, 2, 3, 6, 9, 12\}$. All density forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate the highest ALPL across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 5. Out-of-sample density forecast performance, Large VAR

Variable	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
			$h = 1$					$h = 2$		
PAYEMS	0.038	0.107***	**0.259*****	0.065***	0.063***	0.130***	0.168***	**0.399*****	0.120***	0.113***
CPIAUCSL	-0.401	-0.239	-0.775	-0.078	**0.104**	-0.916	-0.581	-2.312	-0.222	**-0.220**
FEDFUNDS	0.017	0.060***	**0.149*****	-0.017	-0.018	-0.016	**0.021****	-0.023	-0.004	-0.012
INDPRO	-0.050	-0.034	-0.059	-0.011	**0.045*****	0.086	0.105	**0.240****	0.102*	0.146
UNRATE	-1.902	**0.138*****	0.122**	0.096***	0.061***	-1.313	0.144**	**0.235****	0.169***	0.144**
PPIFGS	-0.025	-0.030	-0.711	-0.023	**0.028**	-0.567	-0.213	-1.207	**0.023**	-0.144
GS10	**0.042***	0.036	-0.012	0.001	0.011	-0.011	**0.003**	-0.010	-0.021	-0.027
			$h = 3$					$h = 6$		
PAYEMS	0.141***	0.141***	**0.407*****	0.148***	0.149***	0.116**	0.068***	**0.282*****	0.114***	0.157***
CPIAUCSL	-0.588	**-0.232**	-1.949	-0.446	-0.269	0.010	**0.048**	-0.889	-0.189	-0.002
FEDFUNDS	-0.042	0.027***	**0.032**	0.001	-0.001	-0.019	0.007*	**0.158*****	-0.013	-0.014
INDPRO	0.073*	0.199	0.044	**0.223**	0.074*	**0.060***	-0.053	-0.294	-0.045	-0.081
UNRATE	-1.001	0.116	0.356	**0.357***	0.340*	-0.078	0.495	0.502	**0.958**	0.869
PPIFGS	-0.431	**0.004**	-1.109	-0.020	-0.009	-0.151	-0.108	-0.857	-0.116	**-0.060**
GS10	0.001	**0.002**	-0.025	-0.002	-0.029	-0.006	**0.002**	-0.009	-0.021	-0.024
			$h = 9$					$h = 12$		
PAYEMS	0.067*	0.027*	**0.105***	0.065***	0.099***	**0.062***	0.039	0.016	0.041**	0.024
CPIAUCSL	-0.249	**0.027**	-0.943	-0.081	-0.228	**-0.027**	-0.127	-0.784	-0.106	-0.053
FEDFUNDS	-0.011	-0.006	**0.148*****	-0.024	-0.022	-0.003	-0.005	**0.136*****	-0.028	-0.016
INDPRO	0.105	**0.122**	-0.168	0.092	-0.002	**0.160**	-0.057	-0.231	-0.109	0.058
UNRATE	0.892	**1.495**	0.180	1.326	1.136	1.499	**1.878**	-0.016	1.367	1.040
PPIFGS	-0.165	0.014	-0.629	0.029	**0.061**	**-0.038**	-0.185	-0.711	-0.150	-0.145
GS10	-0.012	-0.011	**0.024**	-0.022	-0.024	-0.034	-0.008	**0.010**	-0.018	-0.040

This table reports the average log predictive likelihood (ALPL) differential between model i and the benchmark AR(1) for the Large size VAR, across a number of different forecast horizons h. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ and $h \in \{1, 2, 3, 6, 9, 12\}$. See notes under Table 4 for additional details.

Table 6. Out-of-sample density forecast performance, Huge VAR

Variable	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
			$h=1$					$h=2$		
PAYEMS	0.199***	0.107***	**0.328*** **	0.094***	0.094***	0.241***	0.206***	**0.498*** **	0.189***	0.193***
CPIAUCSL	-0.098	-0.087	-0.413	-0.018	**0.021**	-0.658	-0.195	-2.113	**0.107*** **	0.028
FEDFUNDS	0.022	0.057***	**0.283*** **	-0.013	-0.010	0.017	0.030	**0.248*** **	-0.006	-0.005
INDPRO	-0.157	**0.052**	-0.389	-0.007	0.002	0.013	0.126***	-0.100	**0.168**	0.126
UNRATE	-1.085	**0.163*** **	0.044	0.106***	0.101***	-0.730	**0.256*** **	0.223**	0.202***	0.190***
PPIFGS	-0.076	0.049	-1.023	**0.083**	0.053	-0.041	-0.002	-1.737	**0.097**	0.092
GS10	-0.008	**0.020**	0.010	-0.004	-0.002	**0.003**	-0.014	-0.010	-0.005	-0.010
			$h=3$					$h=6$		
PAYEMS	0.199***	0.202***	**0.487*** **	0.207***	0.210***	0.072	0.121***	**0.374*** **	0.192***	0.200***
CPIAUCSL	**-0.105**	-0.234	-2.392	-0.209	-0.122	0.000	0.005	-2.254	**0.113*** **	-0.223
FEDFUNDS	-0.005	0.024	**0.229*** **	-0.004	-0.003	0.002	0.012**	**0.190*** **	-0.010	-0.008
INDPRO	-0.005	0.048*	0.122**	**0.147**	0.087*	0.001	-0.019	-0.160	**0.031*** **	-0.046
UNRATE	-0.402	**0.396** **	0.244	0.367**	0.373*	0.535	0.895	-0.666	0.940	**0.987**
PPIFGS	-0.199	-0.029	-1.296	0.021	**0.073**	-0.174	**0.005**	-1.533	-0.002	-0.115
GS10	**0.006**	-0.004	-0.041	0.002	-0.008	**-0.003**	-0.006	-0.024	-0.011	-0.009
			$h=9$					$h=12$		
PAYEMS	0.053	0.081**	**0.249*** **	0.161***	0.166***	0.010	0.035*	**0.146** *	0.091***	0.098***
CPIAUCSL	-0.003	**0.178**	-1.029	-0.076	0.059	-0.013	0.034	-1.486	0.055	**0.162** *
FEDFUNDS	0.011	0.009*	**0.278*** **	-0.013	-0.007	0.002	-0.001	**0.218*** **	-0.021	-0.020
INDPRO	0.020	-0.050	-0.208	**0.052**	0.035*	0.035	**0.075**	-0.214	-0.024	0.003
UNRATE	1.313	**1.476**	-0.519	1.447	1.142	2.235	**2.262**	-1.021	2.095	1.296
PPIFGS	-0.240	**-0.101**	-1.370	-0.264	-0.189	-0.062	0.112	-0.533	**0.180**	0.001
GS10	-0.007	-0.002	**0.032**	-0.012	-0.015	-0.010	0.013	**0.039**	-0.002	-0.016

This table reports the average log predictive likelihood (ALPL) differential between model i and the benchmark AR(1) for the Huge size VAR, across a number of different forecast horizons h. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$ and $h \in \{1, 2, 3, 6, 9, 12\}$. See notes under Table 4 for additional details.

Table 7. Out-of-sample forecast performance: Multivariate results

Fcst h.	*Medium VAR*									
	WTMSFE					*MVALPL*				
	DFM	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
h= 1	1.232	1.143	1.194	**0.936*** **	0.938***	-1.188	0.788***	**1.005*** **	0.919***	0.318***
h= 2	1.092	1.125	1.154	0.937*	**0.936** **	-0.590	0.912***	**1.222*** **	1.120***	0.514***
h= 3	1.066	1.051	1.082	0.949	**0.939* **	-0.277	1.053***	**1.362*** **	1.222***	0.575***
h= 6	1.035	0.961	1.005	**0.936**	0.938	0.255	1.216***	**1.472*** **	1.383***	0.690***
h= 9	1.019	0.934	0.973	**0.926* **	0.930*	0.638	1.336***	**1.502*** **	1.471***	0.703***
h=12	1.017	**0.941**	1.002	0.954	0.960	0.911	1.434**	1.425***	**1.465*** **	0.699***
	Large VAR									
	DFM	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
h= 1	1.288	1.080	1.172	**0.968**	0.975	-1.292	0.830***	**0.913*** **	0.827***	0.257***
h= 2	1.255	1.051	1.224	**0.961**	0.979	-0.685	0.970***	0.937***	**1.071*** **	0.323***
h= 3	1.211	0.969	1.192	**0.962**	0.964	-0.407	1.089***	1.024***	**1.159*** **	0.424***
h= 6	1.110	**0.949** **	0.994	0.954	0.950*	0.210	1.170***	**1.347*** **	1.280***	0.551***
h= 9	1.080	0.967*	0.988	0.953	**0.944* **	0.623	1.311***	1.337***	**1.381*** **	0.574***
h=12	1.061	0.971	1.033	0.980	**0.960**	0.906	**1.390*** **	1.134***	1.362***	0.476***
	Huge VAR									
	DFM	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*	*DFM*	*FAVAR*	*BVAR*	*BCVAR*	*BCVAR$_c$*
h= 1	1.117	1.050	1.055	**0.920*** **	0.944***	-0.342	**0.931*** **	0.760***	0.921***	0.272***
h= 2	1.097	1.023	1.098	**0.916** **	0.923**	0.123	1.148***	0.875***	**1.203*** **	0.468***
h= 3	1.061	0.971	1.061	**0.917* **	0.924*	0.375**	**1.282*** **	1.012***	1.276***	0.510***
h= 6	1.055	0.927*	0.993	0.924	**0.920**	0.813**	1.426***	1.018***	**1.483*** **	0.675***
h= 9	1.028	0.930**	0.962	0.919	**0.915* **	1.108**	1.542***	1.027***	**1.555*** **	0.737***
h=12	1.024	0.955*	0.997	0.949	**0.933**	1.301**	**1.631*** **	0.712	1.593***	0.645***

The left half of this table reports the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model i and the WMSFE of the benchmark AR(1) model, computed as

$$WMSFE_{ih} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{i,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{bcmk,\tau+h}},$$

where $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$ denote the weighted forecast errors of model i and the benchmark model at time $\tau + h$, $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. We set $N = 7$, to focus on the following key seven series, $\{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$. In addition, we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$, and $h \in \{1, 2, 3, 6, 9, 12\}$. The right half of the table shows the multivariate average log predictive likelihood differentials between model i and the benchmark AR(1), computed as

$$MVALPL_{ih} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(MVLPL_{i,\tau+h} - MVLPL_{bcmk,\tau+h}\right),$$

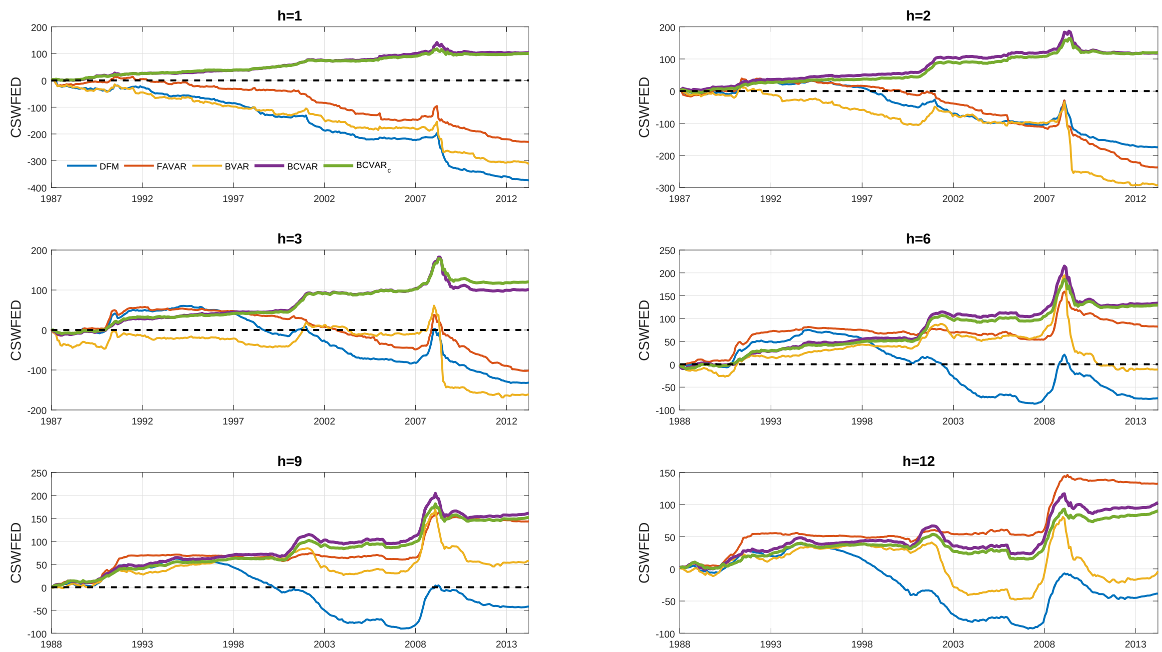
where $MVLPL_{i,\tau+h}$ and $MVLPL_{bcmk,\tau+h}$ denote the multivariate log predictive likelihoods of model i and the benchmark model at time $\tau + h$, and are computed under the assumption of joint normality. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate the lowest WMSFE and highest MVALPL across all models for any given VAR size - forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 8. Out-of-sample forecast performance: Compressed TVP-SV VAR

Variable	*Medium VAR*											
	MSFE						*ALPL*					
	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
PAYEMS	**0.691*****	**0.547*****	**0.525*****	**0.605****	**0.704****	**0.805**	**0.346*****	**0.409*****	**0.374*****	0.155	-0.315	-0.593
CPIAUCSL	**0.912*****	**0.871*****	**0.881*****	**0.857*****	**0.814*****	**0.827*****	**0.254**	**0.317***	**0.286***	**0.229****	**0.283***	**0.275**
FEDFUNDS	**0.888**	**0.932**	**0.931**	**0.958**	0.952	1.022	**0.679*****	**0.557***	**0.490**	**0.315**	0.057	**0.360**
INDPRO	0.912***	0.929	0.940	0.973	0.989	0.974	-0.064	0.067	-0.225	-0.220	-0.239	-0.139
UNRATE	0.807***	0.663***	0.599**	0.561**	0.596**	0.636**	0.132***	0.279***	**0.478****	**0.970**	0.904	1.210
PPIFGS	0.977	**0.984**	**0.989**	**1.014**	0.999	1.019	**0.323**	**0.281**	**0.346**	**0.372**	**0.336**	**0.318**
GS10	1.023	1.022	1.048	1.027	1.020	1.039	0.008	**0.018**	-0.073	**0.040****	-0.046	**0.004**
Multivariate	**0.920*****	**0.894*****	**0.887*****	**0.891****	**0.892****	**0.922***	**1.661*****	**1.791*****	**1.786*****	**1.680*****	1.496***	1.240***
	Large VAR											
	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
PAYEMS	**0.692*****	0.560***	0.573***	**0.627*****	**0.715****	**0.816***	**0.330*****	0.373***	0.243*	-0.007	-0.232	-0.924
CPIAUCSL	**0.941**	**0.879*****	**0.878*****	**0.841*****	**0.811*****	**0.792*****	**0.221**	**0.242*****	**0.186**	**0.257**	**0.265****	**0.252**
FEDFUNDS	**0.903***	**0.861***	**0.888**	**0.922**	**0.968**	1.010	**0.641*****	**0.729*****	**0.683*****	**0.437**	**0.164**	**0.210**
INDPRO	0.922***	0.939*	0.963	**0.933**	**0.967**	0.985	-0.003	0.056	-0.150	-0.315	-0.378	-0.260
UNRATE	0.819***	0.709***	0.670***	0.664***	0.699***	0.741***	0.120***	0.223***	**0.413****	0.732	0.407	-0.956
PPIFGS	**0.969**	**0.984**	**0.991**	**0.991**	**0.980**	0.972	**0.295***	**0.298**	**0.342**	**0.322**	**0.245**	**0.189**
GS10	1.036	1.038	**1.023**	1.035	1.003	1.012	0.015	-0.053	-0.003	-0.040	-0.028	-0.088
Multivariate	**0.930*****	**0.890*****	**0.887*****	**0.880*****	**0.891*****	**0.912****	**1.580*****	**1.718*****	**1.680*****	**1.435*****	1.167***	0.800
	Huge VAR											
	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=9$	$h=12$
PAYEMS	0.713***	0.604***	0.563***	0.669*	0.790	0.880	**0.337*****	0.378***	0.314***	-0.007	-0.572	-0.701
CPIAUCSL	0.972	**0.848*****	**0.876****	**0.853***	**0.835*****	**0.825*****	**0.149**	**0.353***	**0.352**	**0.294**	**0.199**	**0.276**
FEDFUNDS	**0.869***	**0.912**	**0.941**	0.970	1.008	1.102	**0.642****	**0.586***	**0.492**	**0.204**	0.146	0.186
INDPRO	0.937	0.992	0.960	0.989	1.014	1.006	0.000	-0.078	-0.265	-0.379	-0.340	-0.199
UNRATE	0.839**	0.673**	0.590**	0.551**	0.596**	0.638**	0.110***	0.233***	**0.426****	0.879	0.852	0.750
PPIFGS	0.985	1.021	**1.009**	**1.010**	**0.990**	1.024	**0.291***	**0.428**	**0.377**	**0.351**	**0.168**	**0.400**
GS10	1.033	1.044	**1.024**	1.034	1.022	1.033	-0.005	-0.025	-0.013	**0.052****	-0.002	-0.063
Multivariate	0.938***	**0.913****	**0.895****	**0.906**	0.922	0.950	**1.566*****	**1.698*****	**1.754*****	**1.539*****	1.235***	1.041***

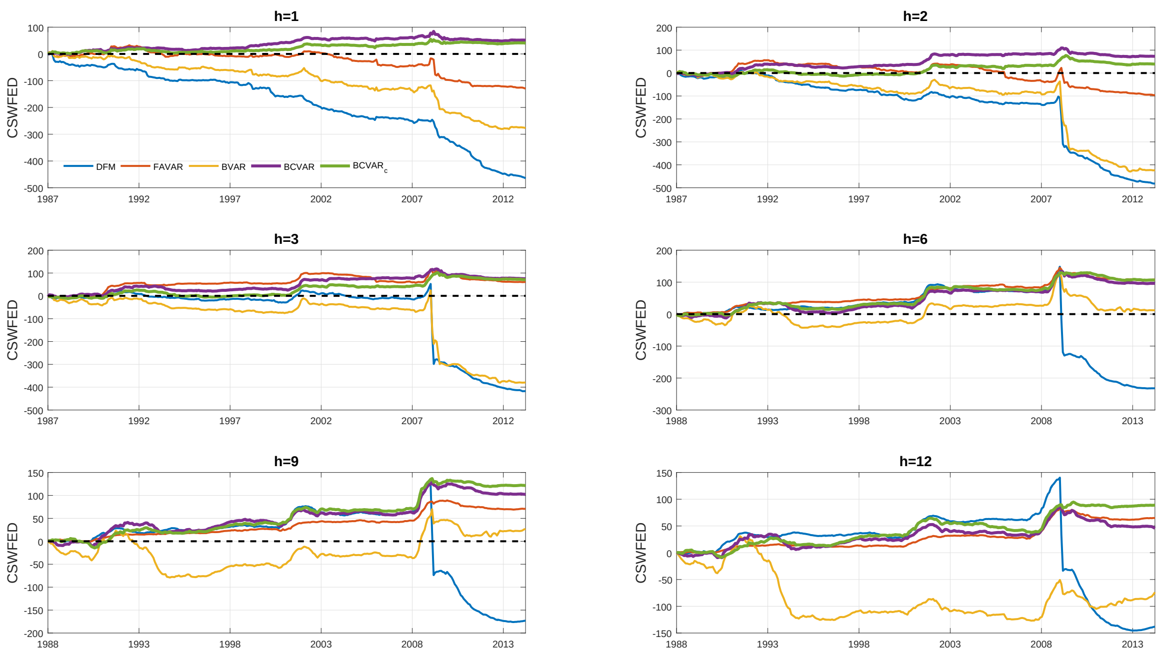
The left half of this table reports the ratio between the univariate or multivariate weighted mean squared forecast error of the $BCVAR_{tvp}$ model and the univariate or multivariate weighted mean squared forecast error of the benchmark AR(1) model. The right half of the table shows the univariate or multivariate average log predictive likelihood differentials between the $BCVAR_{tvp}$ model and the benchmark AR(1) model. h denotes the forecast horizons, with $h \in \{1, 2, 3, 6, 9, 12\}$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Bold numbers indicate all instances where the $BCVAR_{tvp}$ model outperforms all alternative models $(DFM, FAVAR, BVAR, BCVAR, BCVAR_c)$, for any given VAR size/variable/forecast horizon combination. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Figure 1. Cumulative sum of weighted forecast error differentials, Medium VAR



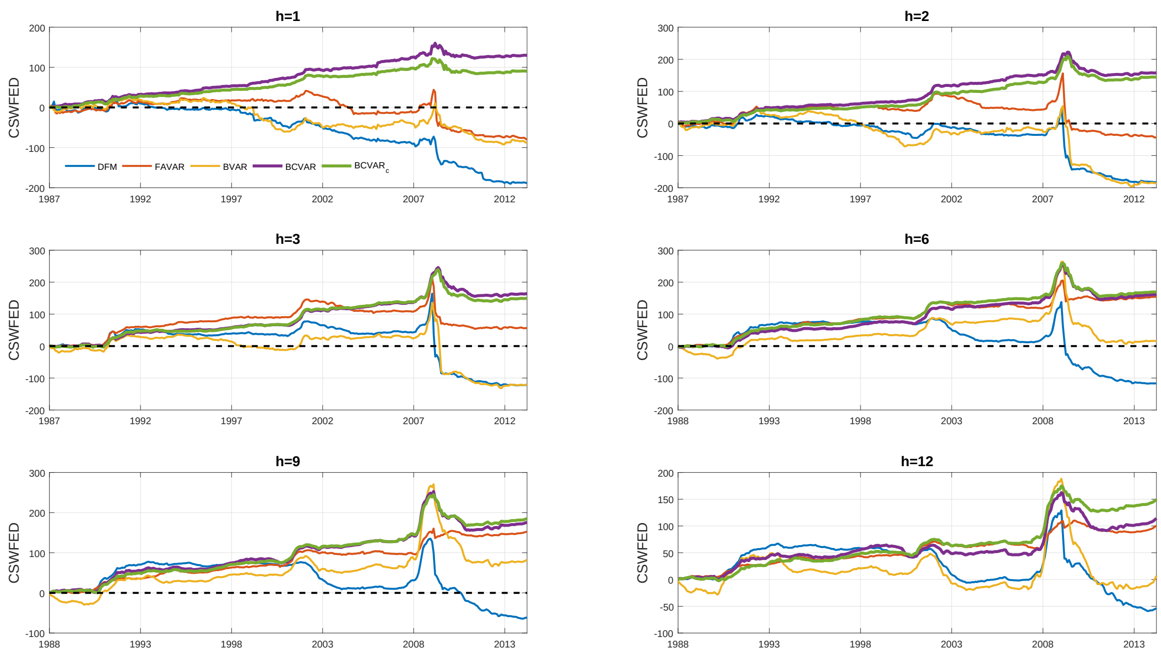
This figure plots the cumulative sum of weighted forecast errors generated by the AR(1) model minus the cumulative sum of weighted forecast errors generated by model i for a Medium size VAR. We define the weighted forecast error of model i and the AR(1) model at time $\tau + h$ as $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$, where $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. We set $N = 7$, to focus on the following key seven series, $\{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$. In addition, we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$, and $h \in \{1, 2, 3, 6, 9, 12\}$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Each panel displays results for a different forecast horizon.

Figure 2. Cumulative sum of weighted forecast error differentials, Large VAR



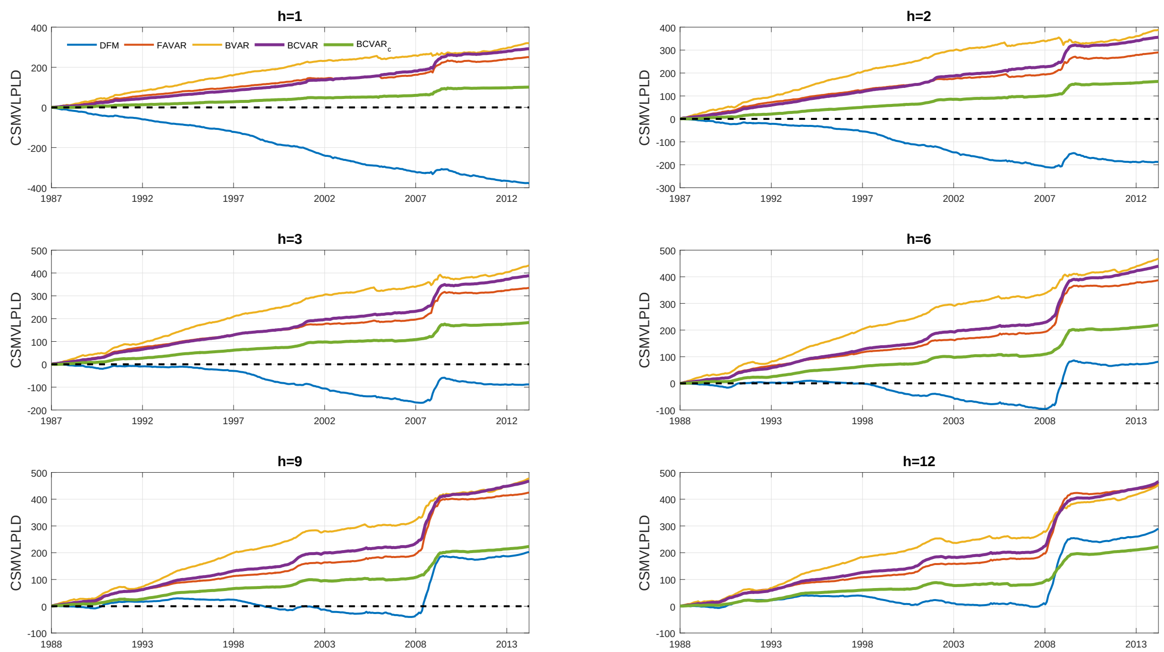
This figure plots the cumulative sum of weighted forecast errors generated by the AR(1) model minus the cumulative sum of weighted forecast errors generated by model i for a Large size VAR. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$. See notes to Figure 1 for additional details.

Figure 3. Cumulative sum of weighted forecast error differentials, Huge VAR



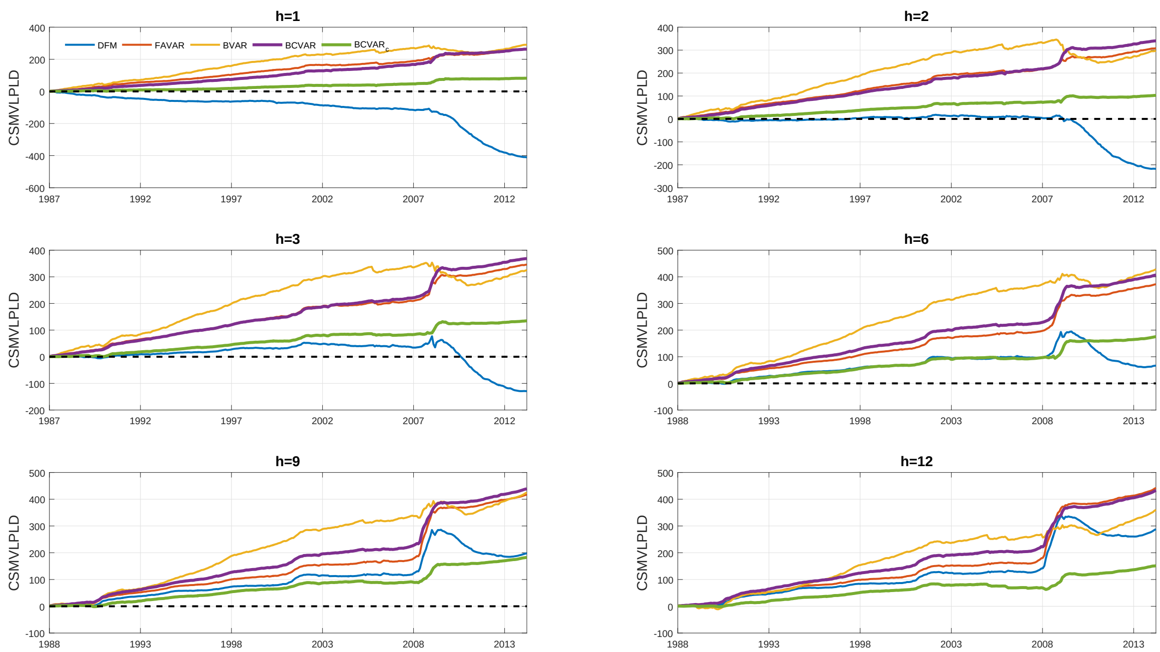
This figure plots the cumulative sum of weighted forecast errors generated by the AR(1) model minus the cumulative sum of weighted forecast errors generated by model i for a Huge size VAR. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$. See notes to Figure 1 for additional details.

Figure 4. Cumulative sum of multivariate log predictive likelihood differentials, Medium VAR



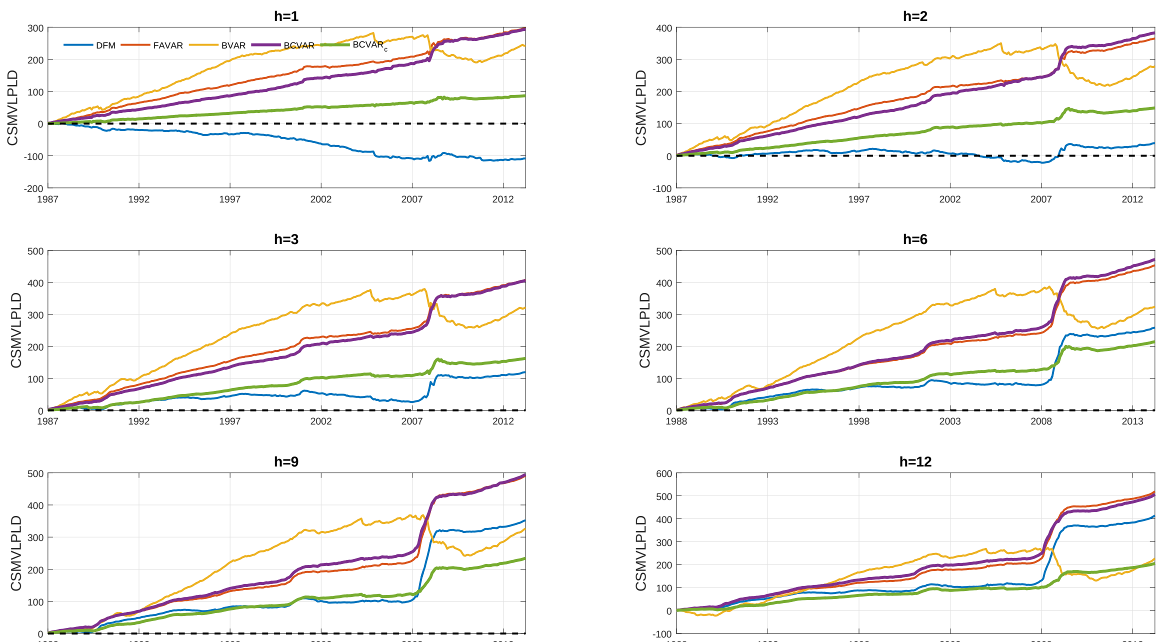
This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by model i minus the cumulative sum of the multivariate log predictive likelihoods computed from an AR(1) model for a Medium size VAR. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$, $h \in \{1, 2, 3, 6, 9, 12\}$, and the multivariate log predictive likelihoods are computed under the assumption of joint normality, as described in the text. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Each panel displays results for a different forecast horizon.

Figure 5. Cumulative sum of multivariate log predictive likelihood differentials, Large VAR



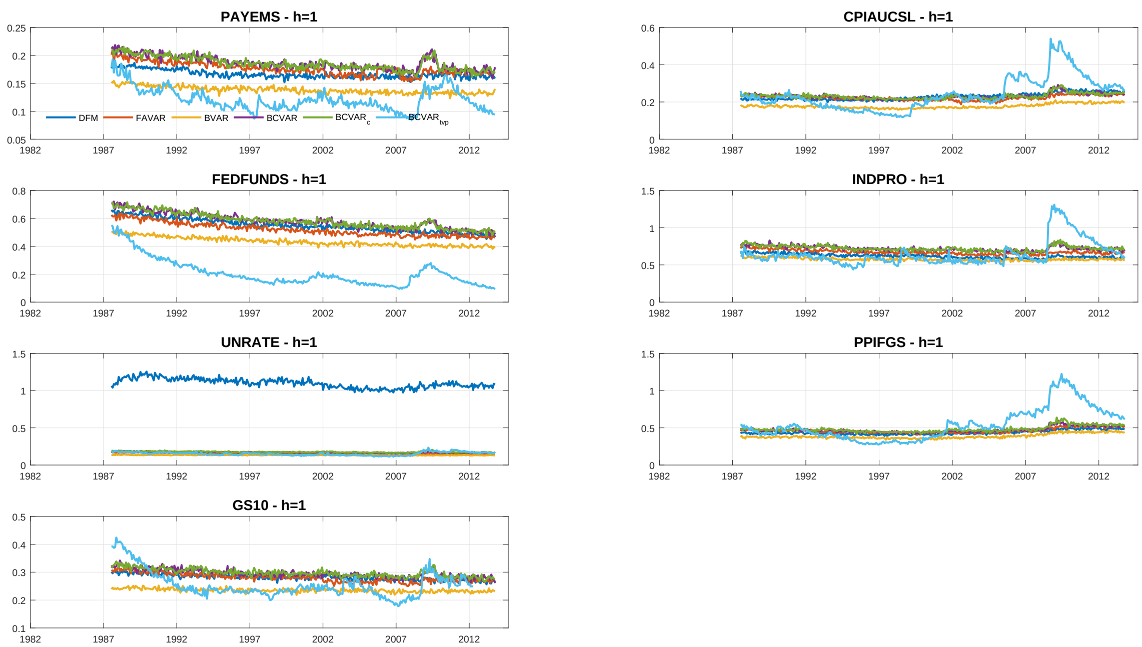
This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by model i minus the cumulative sum of the multivariate log predictive likelihoods computed from an AR(1) model for a Large size VAR. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$. See notes to Figure 4 for additional details.

Figure 6. Cumulative sum of multivariate log predictive likelihood differentials, Huge VAR



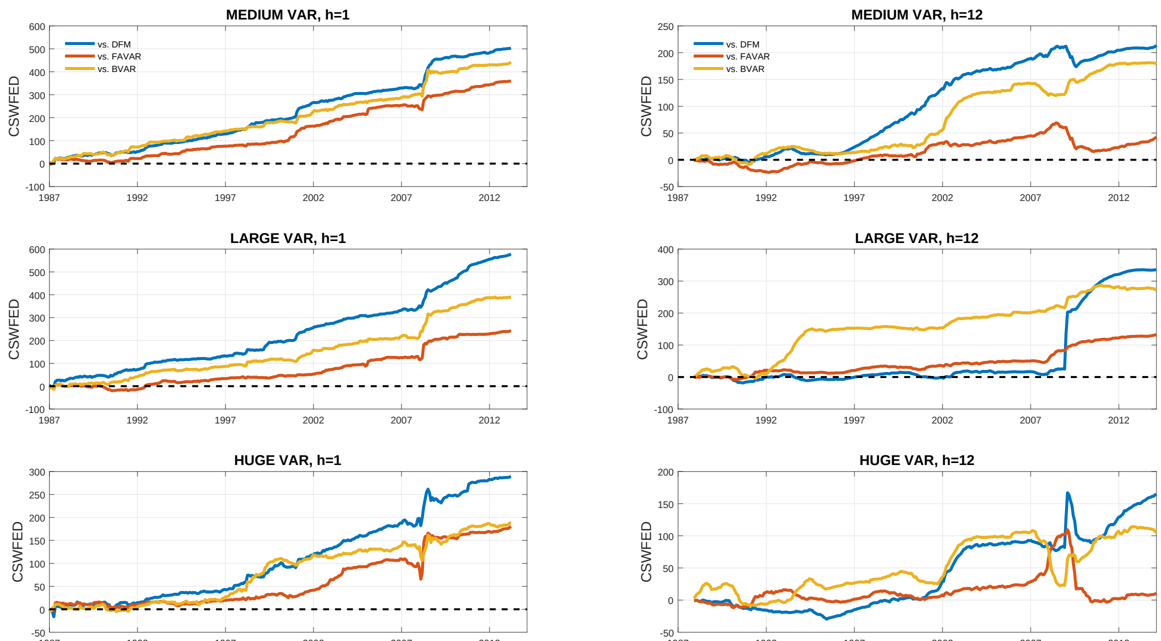
This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by model i minus the cumulative sum of the multivariate log predictive likelihoods computed from an AR(1) model for a Huge size VAR. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c\}$. See notes to Figure 4 for additional details.

Figure 7. Predictive density volatilities, Medium VAR



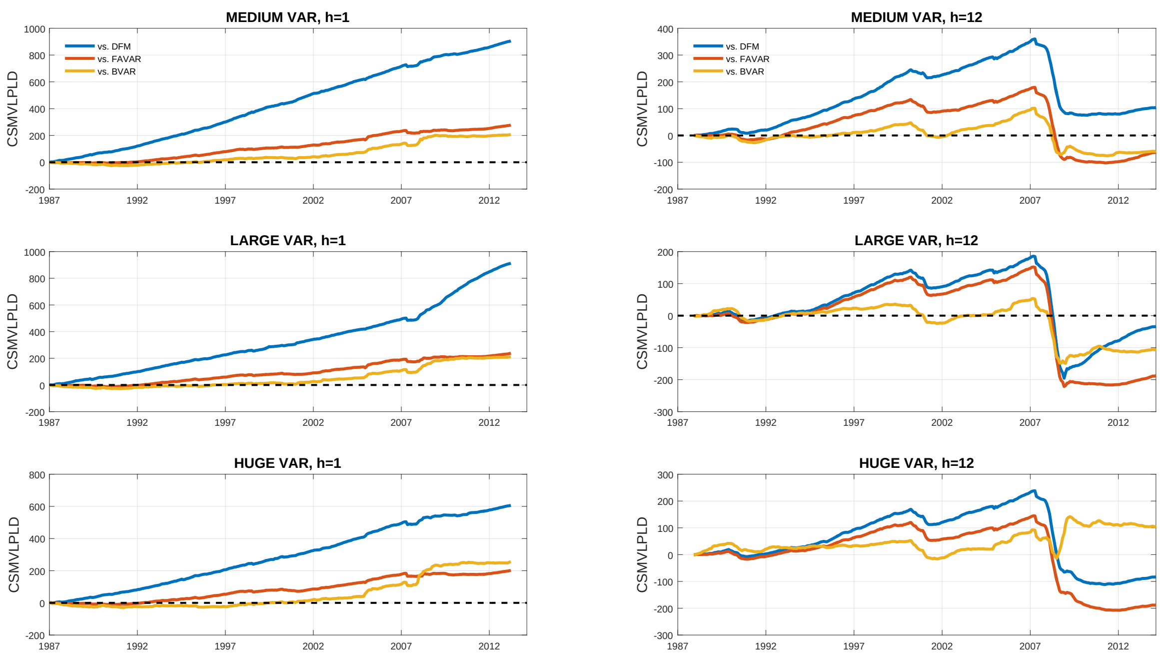
This figure plots the time series of the predicted volatilities over the entire out-of-sample period, for $h = 1$ and the different models entertained, $\{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. The out of sample period starts in 1987:07 and ends in 2014:12. Each panel displays results for a different variable j, where $j \in \{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$.

Figure 8. Cumulative sum of weighted forecast error differentials, Compressed TVP-SV VAR



This figure plots the cumulative sum of weighted forecast errors generated by either the DFM, FAVAR, or BVAR models minus the cumulative sum of weighted forecast errors generated by the BCVAR$_{\text{tvp}}$ model for different VAR sizes and forecast horizons. We define the weighted forecast error of the BCVAR$_{\text{tvp}}$ model and model i alternative at time $\tau + h$ as $we_{\tau+h} = (e'_{\tau+h} \times W \times e_{\tau+h})$ and $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$, where $e_{\tau+h}$ and $e_{i,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. We set $N = 7$, to focus on the following key seven series, $\{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$. In addition, we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. h denotes the forecast horizon, with $h \in \{1, 12\}$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12.

Figure 9. Cumulative sum of multivariate log predictive likelihood differentials, Compressed TVP-SV VAR



This figure plots the cumulative sum of the multivariate log predictive likelihoods generated by the $\text{BCVAR}_{\text{tvp}}$ model minus the cumulative sum of the multivariate log predictive likelihoods computed from either the DFM, FAVAR, or BVAR model for different VAR sizes and forecast horizons. The multivariate log predictive likelihoods are computed under the assumption of joint normality, as described in the text. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12.

Appendix A Data and transformations

The column Tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series description in Global Insight is given in the column GSI:Description.

Table A.1. **Output and Income**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
1	5	X	X	RPI	Real Personal Income	PI
2	5		X	W875RX1	RPI ex. Transfers	PI less transfers
6	5	X	X	INDPRO	IP Index	IP: total
7	5			IPFPNSS	IP: Final Products and Supplies	IP: products
8	5			IPFINAL	IP: Final Products	IP: final prod
9	5			IPCONGD	IP: Consumer Goods	IP: cons gds
10	5			IPDCONGD	IP: Durable Consumer Goods	IP: cons dble
11	5			IPNCONGD	IP: Nondurable Consumer Goods	IP: cons nondble
12	5			IPBUSEQ	IP: Business Equipment	IP: bus eqpt
13	5			IPMAT	IP: Materials	IP: matls
14	5			IPDMAT	IP: Durable Materials	IP: dble matls
15	5			IPNMAT	IP: Nondurable Materials	IP: nondble matls
16	5			IPMANSICS	IP: Manufacturing	IP: mfg
17	5			IPB51222S	IP: Residential Utilities	IP: res util
18	5			IPFUELS	IP: Fuels	IP: fuels
19	1		X	NAPMPI	ISM Manufacturing: Production	NAPM prodn
20	1			CAPUTLB00004S	Capacity Utilization: Manufacturing	Cap util

Table A.2. **Labor Market**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
21	1	X	X	HWI	Help-Wanted Index for US	Help wanted indx
22	1		X	HWIURATIO	Help Wanted to Unemployed ratio	Help wanted/unemp
23	5		X	CLF16OV	Civilian Labor Force	Emp CPS total
24	5			CE16OV	Civilian Employment	Emp CPS nonag
25	1	X	X	UNRATE	Civilian Unemployment Rate	U: all
26	1			UEMPMEAN	Average Duration of Unemployment	U: mean duration
27	5			UEMPLT5	Civilians Unemployed $\leq$ 5 Weeks	U $\leq$ 5 wks
28	5			UEMP5TO14	Civilians Unemployed 5-14 Weeks	U 5-14 wks
29	5			UEMP15OV	Civilians Unemployed $>$ 15 Weeks	U $>$ 15 wks
30	5			UEMP15T26	Civilians Unemployed 15-26 Weeks	U 15-26 wks
31	5			UEMP27OV	Civilians Unemployed $>$ 27 Weeks	U $>$ 27 wks
32	5			CLAIMSx	Initial Claims	UI claims
33	5	X	X	PAYEMS	All Employees: Total nonfarm	Emp: total
34	5			USGOOD	All Employees: Goods-Producing	Emp: gds prod
35	5			CES1021000001	All Employees: Mining and Logging	Emp: mining
36	5			USCONS	All Employees: Construction	Emp: const
37	5			MANEMP	All Employees: Manufacturing	Emp: mfg
38	5			DMANEMP	All Employees: Durable goods	Emp: dble gds
39	5			NDMANEMP	All Employees: Nondurable goods	Emp: nondbles
40	5			SRVPRD	All Employees: Service Industries	Emp: services
41	5			USTPU	All Employees: TT&U	Emp: TTU
42	5			USWTRADE	All Employees: Wholesale Trade	Emp: wholesale
43	5			USTRADE	All Employees: Retail Trade	Emp: retail
44	5			USFIRE	All Employees: Financial Activities	Emp: FIRE
45	5			USGOVT	All Employees: Government	Emp: Govt
46	1		X	CES0600000007	Hours: Goods-Producing	Avg hrs
47	1			AWOTMAN	Overtime Hours: Manufacturing	Overtime: mfg
48	1			AWHMAN	Hours: Manufacturing	Avg hrs: mfg
49	1			NAPMEI	ISM Manufacturing: Employment	NAPM empl
128	5			CES0600000008	Ave. Hourly Earnings: Goods	AHE: goods
129	5			CES2000000008	Ave. Hourly Earnings: Construction	AHE: const
130	5			CES3000000008	Ave. Hourly Earnings: Manufacturing	AHE: mfg

Table A.3. **Housing**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
50	4	X		HOUST	Starts: Total	Starts: nonfarm
51	4			HOUSTNE	Starts: Northeast	Starts: NE
52	4			HOUSTMW	Starts: Midwest	Starts: MW
53	4			HOUSTS	Starts: South	Starts: South
54	4			HOUSTW	Starts: West	Starts: West
55	4	X		PERMIT	Permits	BP: total
56	4			PERMITNE	Permits: Northeast	BP: NE
57	4			PERMITMW	Permits: Midwest	BP: MW
58	4			PERMITS	Permits: South	BP: South
59	4			PERMITW	Permits: West	BP: West

Table A.4. **Consumption, Orders and Inventories**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
3	5		X	DPCERA3M086SBEA	Real PCE	Real Consumption
4	5		X	CMRMTSPLx	Real M&T Sales	M&T sales
5	5		X	RETAILx	Retail and Food Services Sales	Retail sales
60	1	X		NAPM	ISM: PMI Composite Index	PMI
61	1		X	NAPMNOI	ISM: New Orders Index	NAPM new ordrs
62	1		X	NAPMSDI	ISM: Supplier Deliveries Index	NAPM vendor del
63	1		X	NAPMII	ISM: Inventories Index	NAPM Invent
65	5			AMDMNOx	Orders: Durable Goods	Orders: dble gds
67	5			AMDMUOx	Unfilled Orders: Durable Goods	Unf orders: dble
68	5			BUSINVx	Total Business Inventories	M&T invent
69	1			ISRATIOx	Inventories to Sales Ratio	M&T invent/sales

Table A.5. **Money and Credit**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
70	5	X	X	M1SL	M1 Money Stock	M1
71	5		X	M2SL	M2 Money Stock	M2
73	5		X	M2REAL	Real M2 Money Stock	M2 (real)
74	5		X	AMBSL	St. Louis Adjusted Monetary Base	MB
75	5		X	TOTRESNS	Total Reserves	Reserves tot
77	5	X	X	BUSLOANS	Commercial and Industrial Loans	C&I loan plus
78	5			REALLN	Real Estate Loans	DC&I loans
79	5		X	NONREVSL	Total Nonrevolving Credit	Cons credit
80	1		X	CONSPI	Credit to PI ratio	Inst cred/PI
132	5			MZMSL	MZM Money Stock	N.A.
133	5			DTCOLNVHFNM	Consumer Motor Vehicle Loans	N.A.
134	5			DTCTHFNM	Total Consumer Loans and Leases	N.A.
135	5	X		INVEST	Securities in Bank Credit	N.A.

Table A.6. **Interest rates and Exchange rates**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
85	2	X	X	FEDFUNDS	Effective Federal Funds Rate	Fed Funds
86	2		X	CP3M	3-Month AA Comm. Paper Rate	Comm paper
87	2		X	TB3MS	3-Month T-bill	3 mo T-bill
88	2		X	TB6MS	6-Month T-bill	6 mo T-bill
89	2		X	GS1	1-Year T-bond	1 yr T-bond
90	2		X	GS5	5-Year T-bond	5 yr T-bond
91	2	X	X	GS10	10-Year T-bond	10 yr T-bond
92	2		X	AAA	Aaa Corporate Bond Yield	Aaa bond
93	2		X	BAA	Baa Corporate Bond Yield	Baa bond
94	1			COMPAPFF	CP - FFR spread	CP-FF spread
95	1			TB3SMFFM	3 Mo. - FFR spread	3 mo-FF spread
96	1			TB6SMFFM	6 Mo. - FFR spread	6 mo-FF spread
97	1			T1YFFM	1 yr. - FFR spread	1 yr-FF spread
98	1			T5YFFM	5 yr. - FFR spread	5 yr-FF spread
99	1	X		T10YFFM	10 yr. - FFR spread	10 yr-FF spread
100	1			AAAFFM	Aaa - FFR spread	Aaa-FF spread
101	1			BAAFFM	Baa - FFR spread	Baa-FF spread
103	5		X	EXSZUS	Switzerland / U.S. FX Rate	Ex rate: Switz
104	5		X	EXJPUS	Japan / U.S. FX Rate	Ex rate: Japan
105	5	X	X	EXUSUK	U.S. / U.K. FX Rate	Ex rate: UK
106	5		X	EXCAUS	Canada / U.S. FX Rate	EX rate: Canada

Table A.7. **Prices**

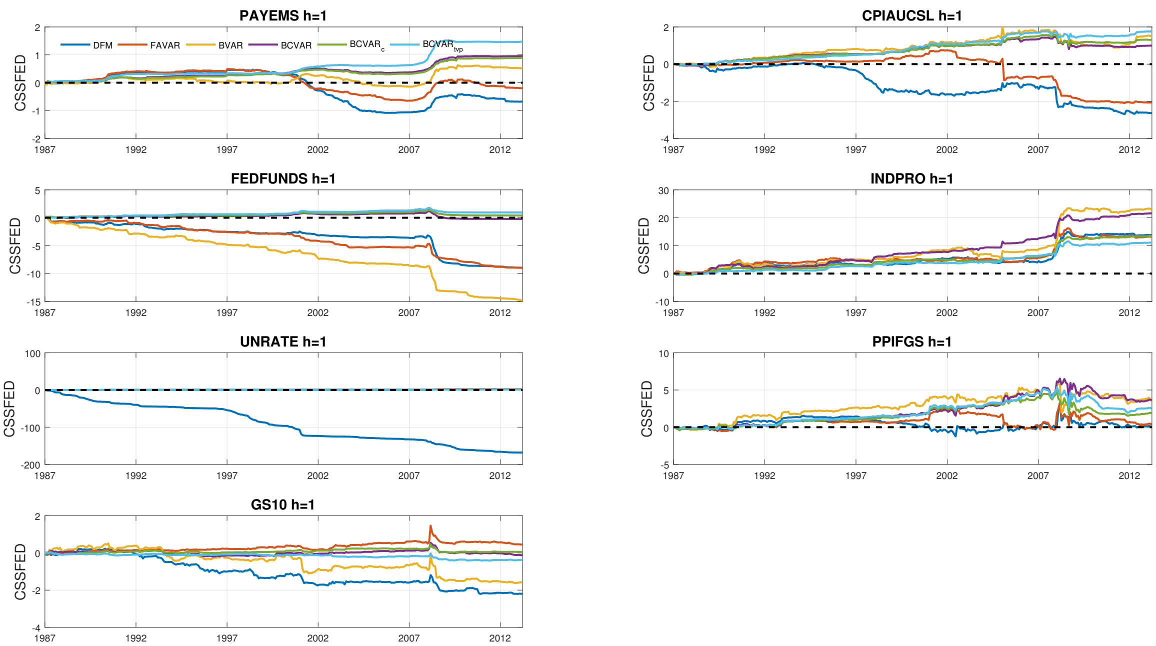
Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
107	5	X	X	PPIFGS	PPI: Finished Goods	PPI: fin gds
108	5		X	PPIFCG	PPI: Finished Consumer Goods	PPI: cons gds
109	5		X	PPIITM	PPI: Intermediate Materials	PPI: int materials
110	5		X	PPICRM	PPI: Crude Materials	PPI: crude materials
111	5	X		oilprice	Crude Oil Prices: WTI	Spot market price
112	5			PPICMM	PPI: Commodities	PPI: nonferrous
113	1			NAPMPRI	ISM Manufacturing: Prices	NAPM com price
114	5	X	X	CPIAUCSL	CPI: All Items	CPI-U: all
115	5			CPIAPPSL	CPI: Apparel	CPI-U: apparel
116	5			CPITRNSL	CPI: Transportation	CPI-U: transp
117	5			CPIMEDSL	CPI: Medical Care	CPI-U: medical
118	5			CUSR0000SAC	CPI: Commodities	CPI-U: comm.
119	5			CUUR0000SAD	CPI: Durables	CPI-U: dbles
120	5			CUSR0000SAS	CPI: Services	CPI-U: services
121	5			CPIULFSL	CPI: All Items Less Food	CPI-U: ex food
122	5			CUUR0000SA0L2	CPI: All items less shelter	CPI-U: ex shelter
123	5			CUSR0000SA0L5	CPI: All items less medical care	CPI-U: ex med
124	5			PCEPI	PCE: Chain-type Price Index	PCE defl
125	5			DDURRG3M086SBEA	PCE: Durable goods	PCE defl: dlbes
126	5			DNDGRG3M086SBEA	PCE: Nondurable goods	PCE defl: nondble
127	5			DSERRG3M086SBEA	PCE: Services	PCE defl: service

Table A.8. **Stock Market**

Series id	*Tcode*	*Medium*	*Large*	*FRED*	*Description*	*GSI:Description*
81	5	X	X	S&P 500	S&P: Composite	S&P 500
82	5		X	S&P: indust	S&P: Industrials	S&P: indust
83	1		X	S&P div yield	S&P: Dividend Yield	S&P div yield
84	5		X	S&P PE ratio	S&P: Price-Earnings Ratio	S&P PE ratio

Appendix B Cumulative sums of squared forecast error and log predictive likelihood differentials for individual series

Figure B.1. Cumulative sum of squared forecast error differentials, Medium VAR, $h = 1$

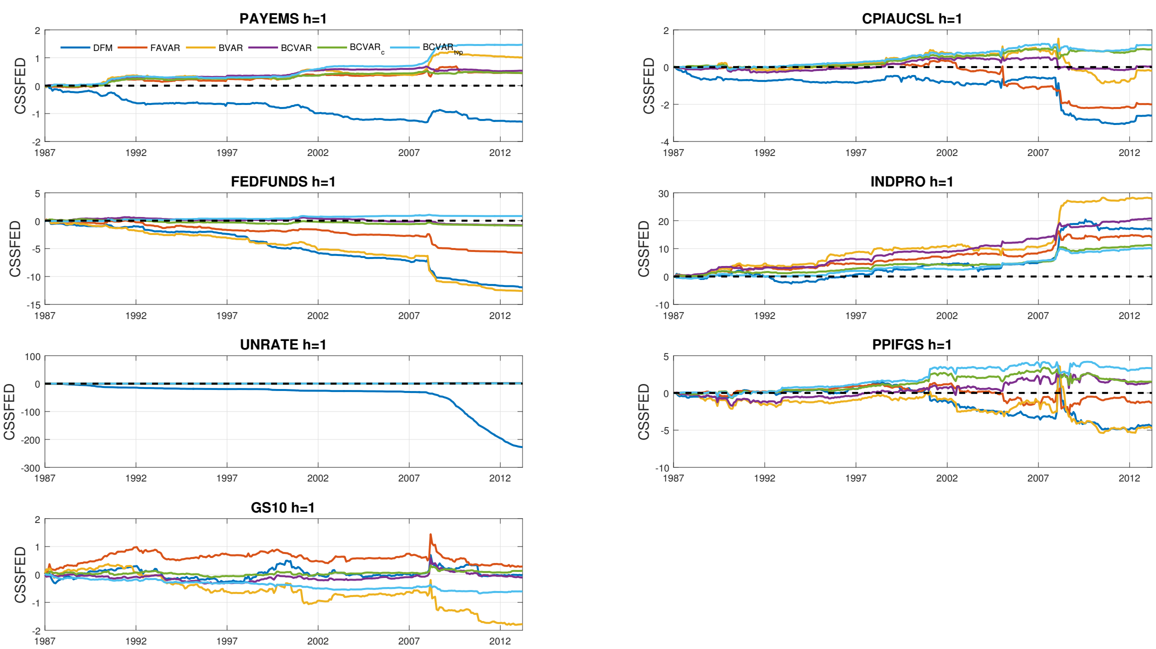


This figure plots the cumulative sum of squared forecast errors generated by the AR(1) model minus the cumulative sum of squared forecast errors generated by model i for a Medium size VAR and forecast horizon $h = 1$,

$$CSSED_{ijht} = \sum_{\tau=\underline{t}}^{t} \left(e^2_{bcmk,j,\tau+h} - e^2_{i,j,\tau+h}\right)$$

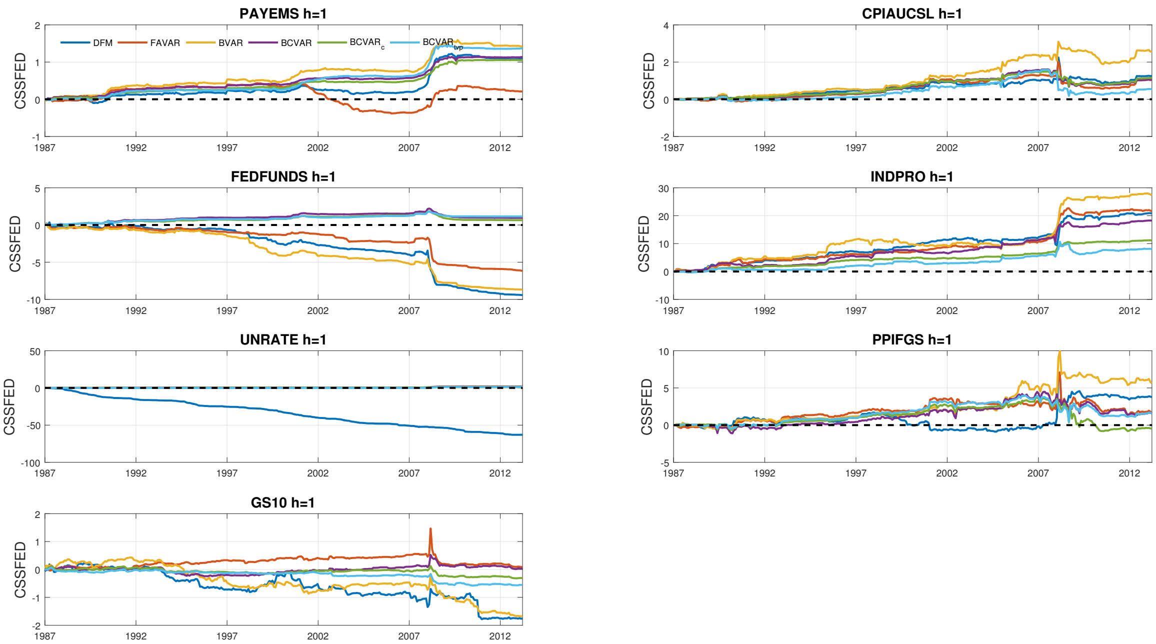
where $t = \underline{t}, ..., \overline{t}-h$. Values above zero indicate that model i generates better performance than the benchmark, while negative values suggest the opposite. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$, $j \in \{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$, $\underline{t}$ and $\overline{t}$ denote the start and end of the out-of-sample period. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Each panel displays results for a different series.

Figure B.2. Cumulative sum of squared forecast error differentials, Large VAR, $h = 1$



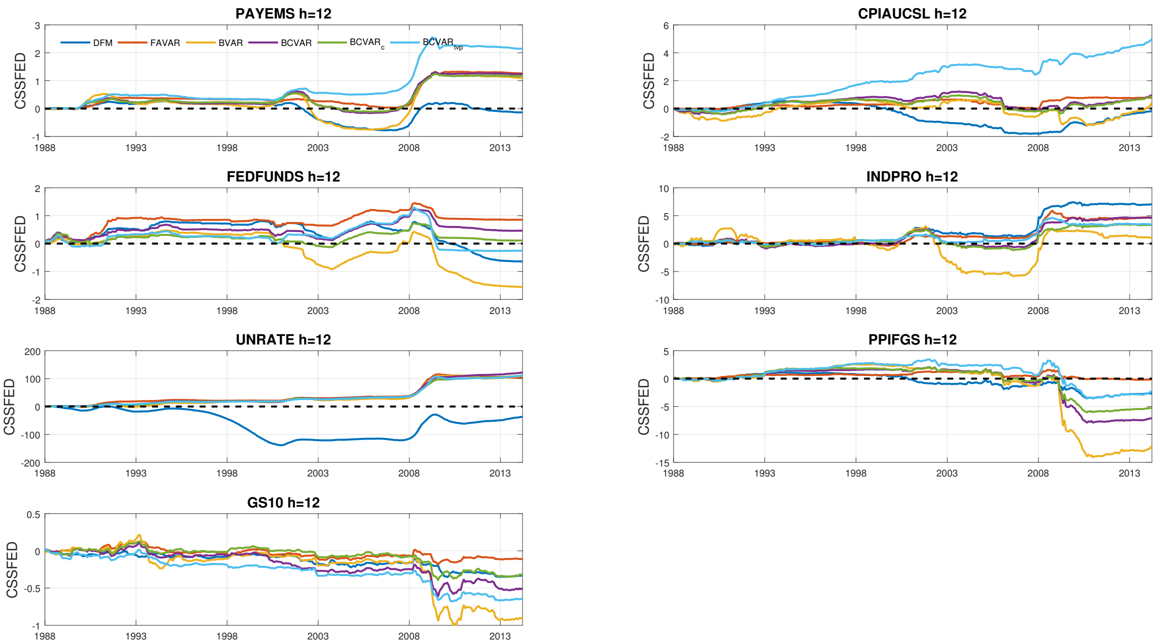
This figure plots the cumulative sum of squared forecast errors generated by the AR(1) model minus the cumulative sum of squared forecast errors generated by model i for a Large size VAR and forecast horizon $h = 1$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.1 for additional details.

Figure B.3. Cumulative sm of squared forecast error differentials, Huge VAR, $h = 1$



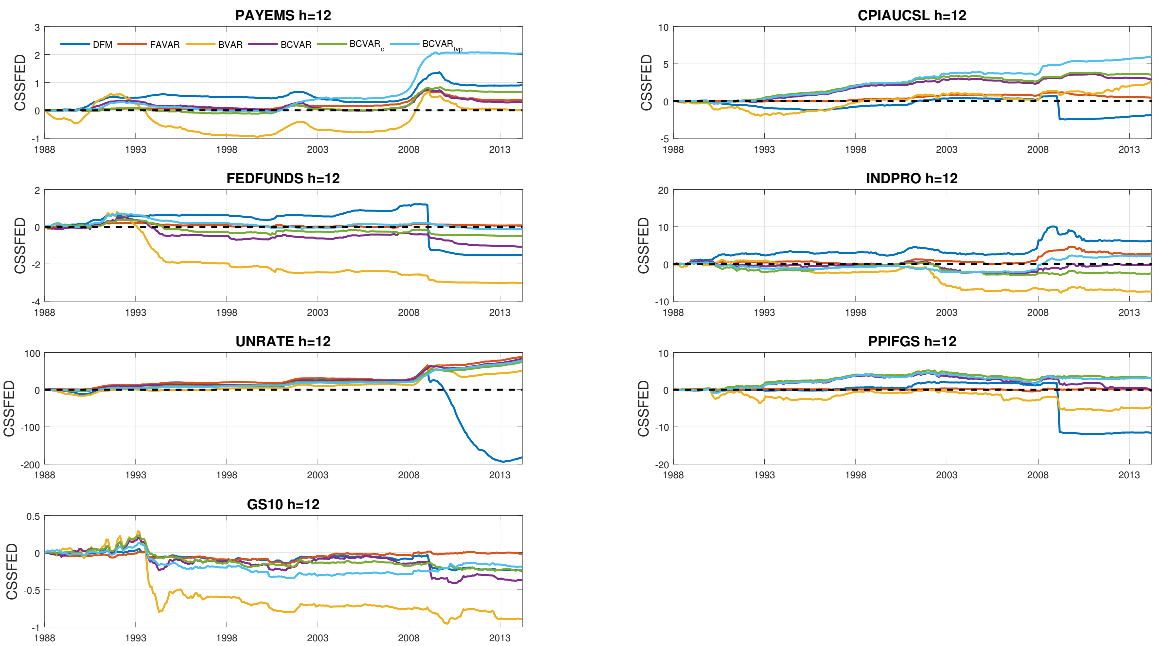
This figure plots the cumulative sum of squared forecast errors generated by the AR(1) model minus the cumulative sum of squared forecast errors generated by model i for a Huge size VAR and forecast horizon $h = 1$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.1 for additional details.

Figure B.4. Cumulative sum of squared forecast error differentials, Medium VAR, $h = 12$



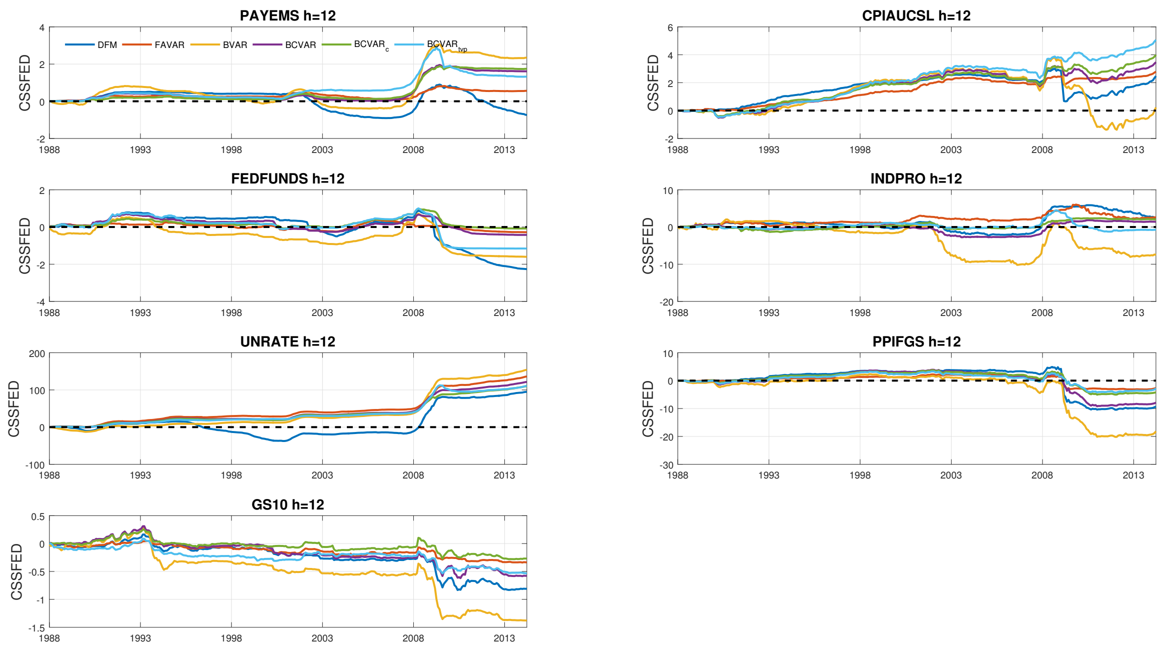
This figure plots the cumulative sum of squared forecast errors generated by the AR(1) model minus the cumulative sum of squared forecast errors generated by model i for a Medium size VAR and forecast horizon $h = 12$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.1 for additional details.

Figure B.5. Cumulative sum of squared forecast error differentials, Large VAR, $h = 12$



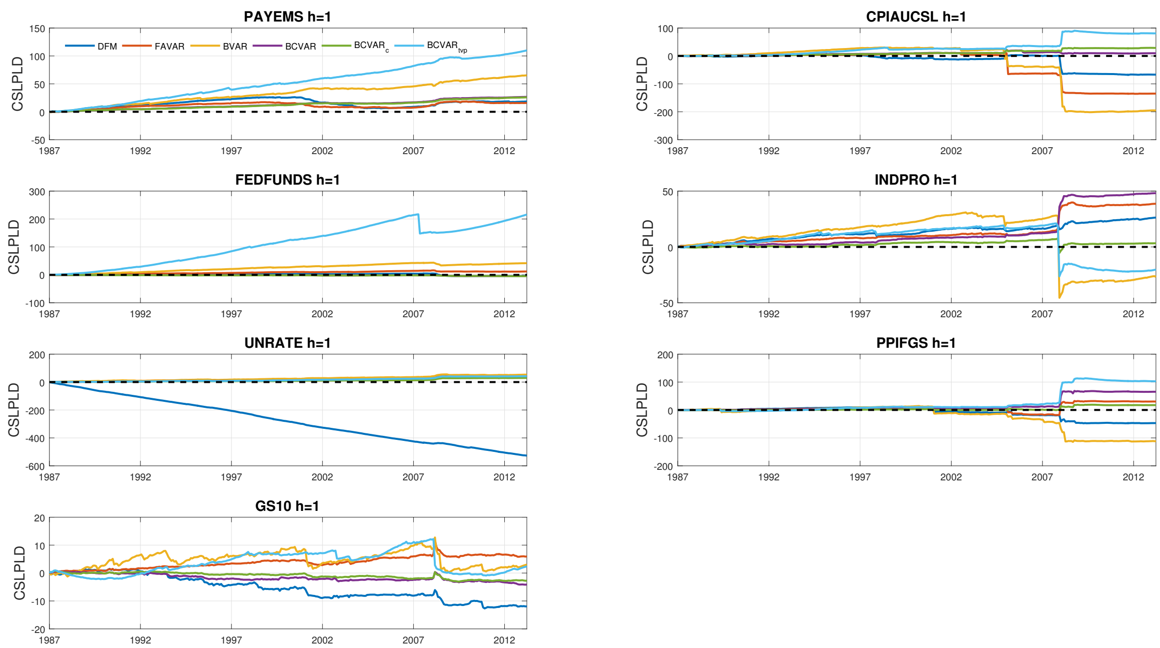
This figure plots the cumulative sum of squared forecast errors generated by the AR(1) model minus the cumulative sum of squared forecast errors generated by model i for a Large size VAR and forecast horizon $h = 12$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.1 for additional details.

Figure B.6. Cumulative sum of squared forecast error differentials, Huge VAR, $h = 12$



This figure plots the cumulative sum of squared forecast errors generated by the AR(1) model minus the cumulative sum of squared forecast errors generated by model i for a Huge size VAR and forecast horizon $h = 12$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.1 for additional details.

Figure B.7. Cumulative sum of log predictive likelihood differentials, Medium VAR, $h = 1$

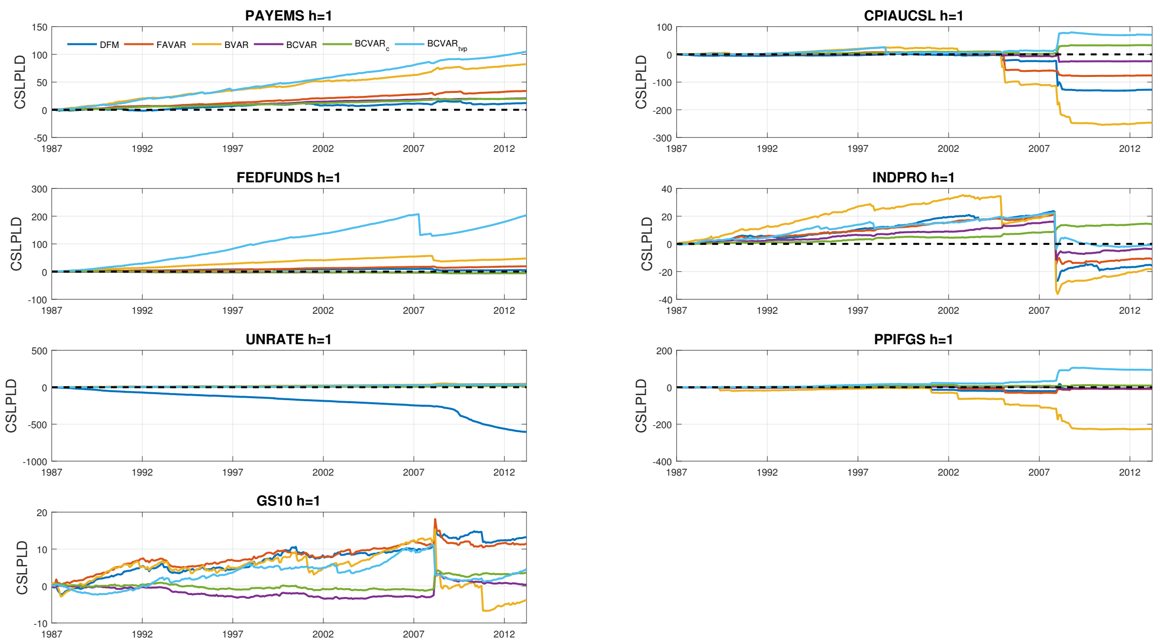


This figure plots the cumulative sum of log predictive likelihoods generated by model i minus the cumulative sum of log predictive likelihoods generated by the AR(1) model for a Medium size VAR and forecast horizon $h = 1$,

$$CLPLD_{ijht} = \sum_{\tau=\underline{t}}^{t} (LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h})$$

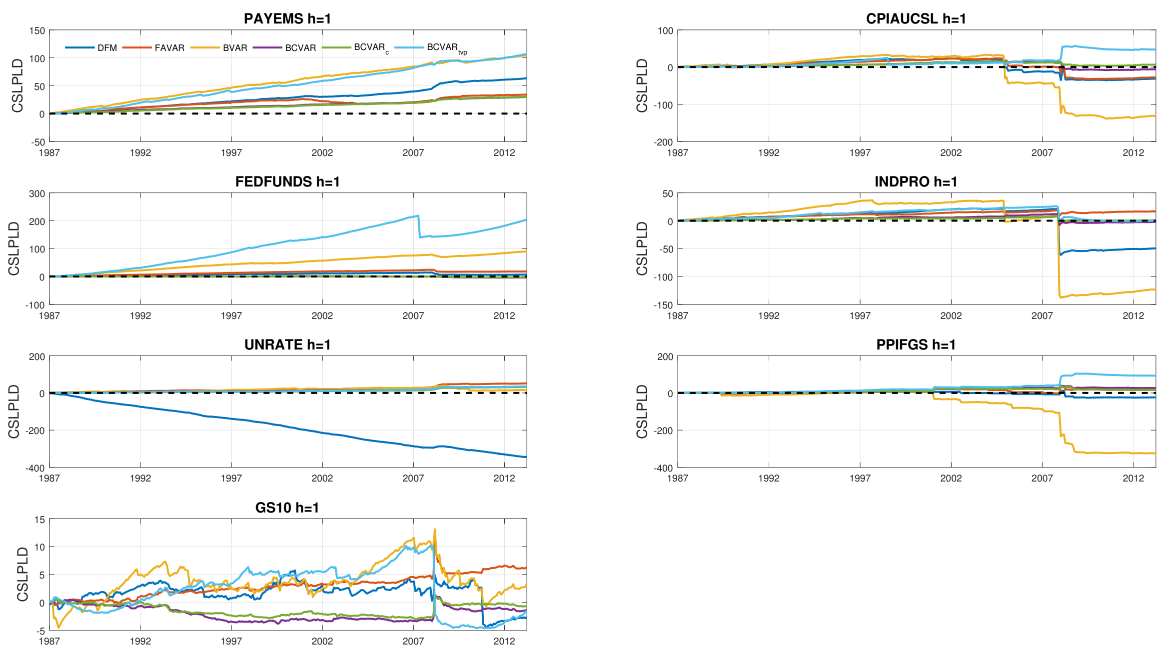
where $t = \underline{t}, ..., \bar{t} - h$. Values above zero indicate that model i generates better performance than the benchmark, while negative values suggest the opposite. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$, $j \in \{PAYEMS, CPIAUCSL, FEDFUNDS, INDPRO, UNRATE, PPIFGS, GS10\}$, $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1987:07 and ending in 2014:12. Each panel displays results for a different series.

Figure B.8. Cumulative sum of log predictive likelihood differentials, Large VAR, $h = 1$



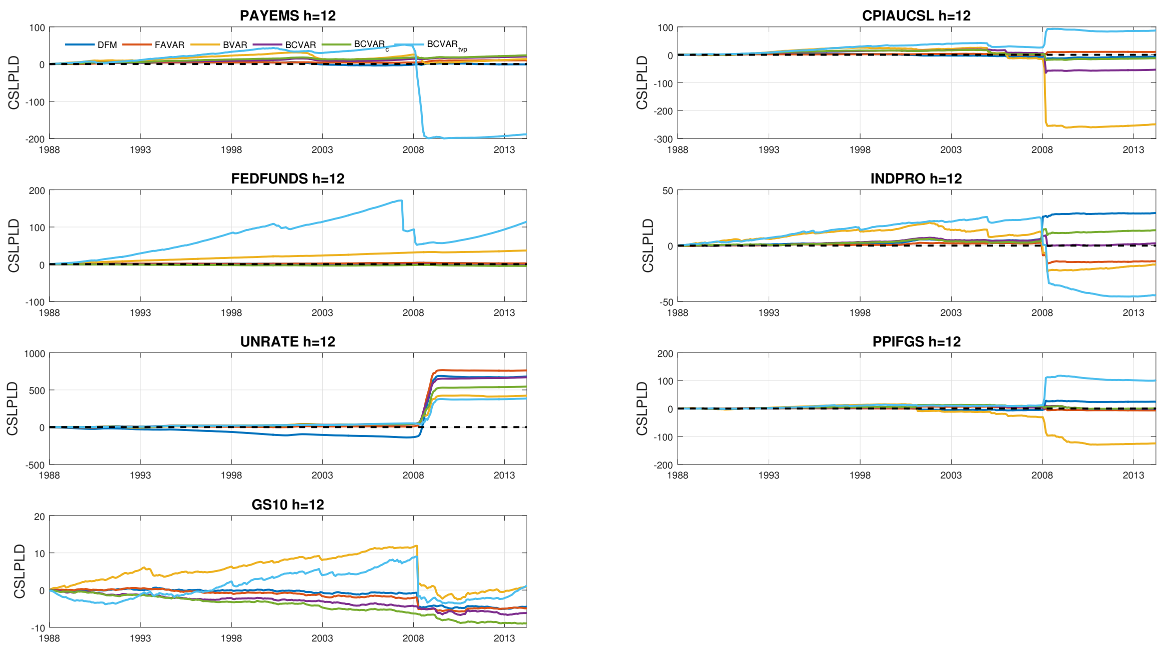
This figure plots the cumulative sum of log predictive likelihoods generated by model i minus the cumulative sum of log predictive likelihoods generated by the AR(1) model for a Large size VAR and forecast horizon $h = 1$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.7 for additional details.

Figure B.9. Cumulative sum of log predictive likelihood differentials, Huge VAR, $h = 1$



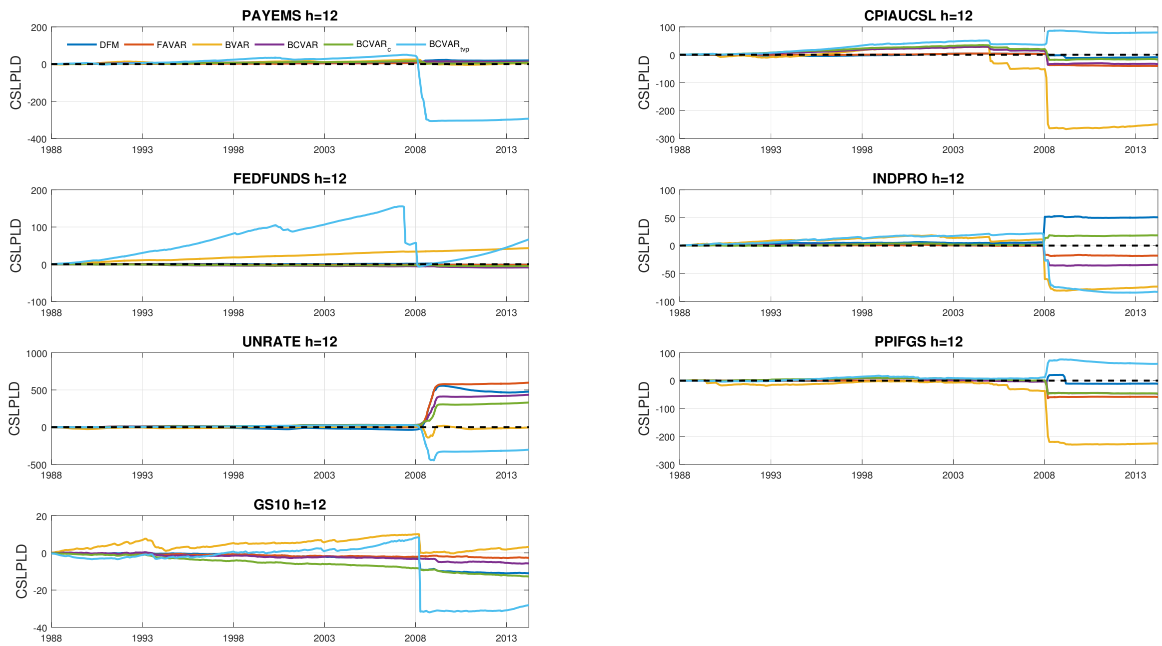
This figure plots the cumulative sum of log predictive likelihoods generated by model i minus the cumulative sum of log predictive likelihoods generated by the AR(1) model for a Huge size VAR and forecast horizon $h = 1$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.7 for additional details.

Figure B.10. Cumulative sum of log predictive likelihood differentials, Medium VAR, $h = 12$



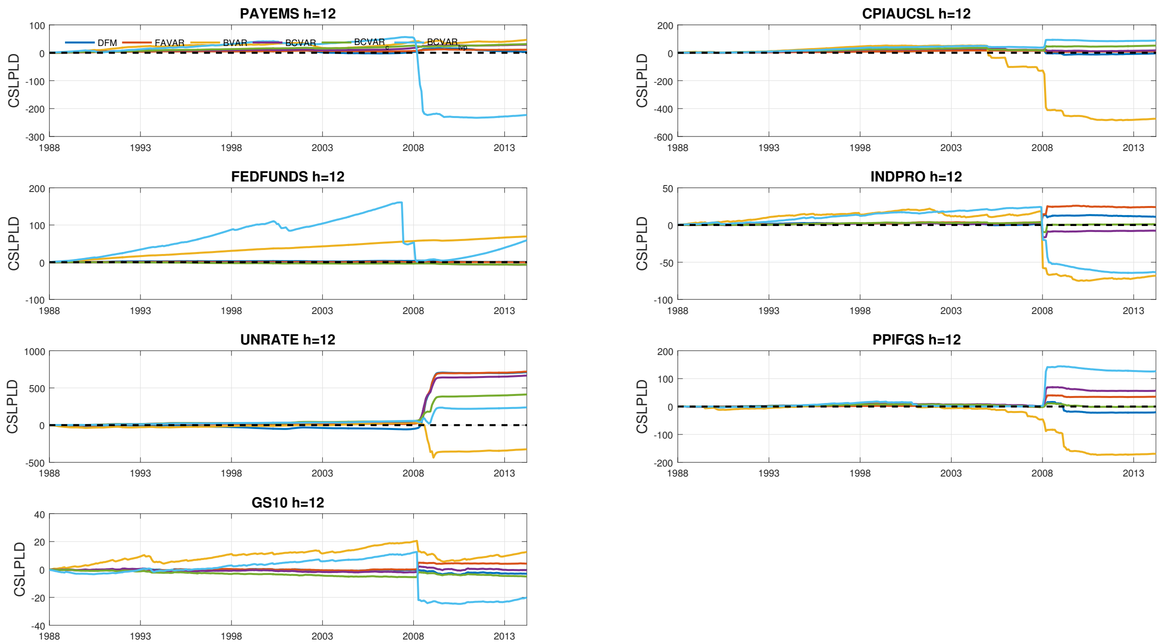
This figure plots the cumulative sum of log predictive likelihoods generated by model i minus the cumulative sum of log predictive likelihoods generated by the AR(1) model for a Medium size VAR and forecast horizon $h = 12$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.7 for additional details.

Figure B.11. Cumulative sum of log predictive likelihood differentials, Large VAR, $h = 12$



This figure plots the cumulative sum of log predictive likelihoods generated by model i minus the cumulative sum of log predictive likelihoods generated by the AR(1) model for a Large size VAR and forecast horizon $h = 12$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.7 for additional details.

Figure B.12. Cumulative sum of log predictive likelihood differentials, Huge VAR, $h = 12$



This figure plots the cumulative sum of log predictive likelihoods generated by model i minus the cumulative sum of log predictive likelihoods generated by the AR(1) model for a Huge size VAR and forecast horizon $h = 12$. $i \in \{DFM, FAVAR, BVAR, BCVAR, BCVAR_c, BCVAR_{tvp}\}$. See notes to Figure B.7 for additional details.