Forecasting in vector autoregressions with many predictors

Korobilis, Dimitris

University of Strathclyde

January 2008

Forecasting in vector autoregressions with many predictors

Dimitris Korobilis*
University of Strathclyde

Working Paper, January 2008

Abstract

This paper addresses the issue of improving the forecasting performance of vector autoregressions (VARs) when the set of available predictors is inconveniently large to handle with methods and diagnostics used in traditional small-scale models. First, available information from a large dataset is summarized into a considerably smaller set of variables through factors estimated using standard principal components. However, even in the case of reducing the dimension of the data the true number of factors may still be large. For that reason I introduce in my analysis simple and efficient Bayesian model selection methods. Model estimation and selection of predictors is carried out automatically through a stochastic search variable selection (SSVS) algorithm which requires minimal input by the user. I apply these methods to forecast 8 main U.S. macroeconomic variables using 124 potential predictors. I find improved out of sample fit in high dimensional specifications that would otherwise suffer from the proliferation of parameters.

Keywords: Bayesian VAR, forecasting, model selection & averaging, large datasets

JEL Classification: C11, C32, C52, C53

*Contact address: Department of Economics, University of Strathclyde, Sir William Duncan Building, 130 Rottenrow, G4 0GE, Glasgow, United Kingdom. Tel: (+44) 0141 548 4340, e-mail:dikorobilis@googlemail.com.

1. INTRODUCTION

It is common practice today to collect observations on many variables that potentially help explain economic variables of interest such as inflation and unemployment. Technological progress has allowed the collection, storage, and exchange of huge amounts of information without much effort and cost. In turn, this has significantly affected recent macroeconomic modeling techniques. Current academic research is focused on finding solutions on how to efficiently handle large amounts of information with, for example, Stock and Watson (2002) using 215 predictors to forecast 8 major macroeconomic variables for the U.S. economy. Bernanke and Boivin (2003), among others, argue that this is also the case nowadays in central banks, where it is customary for researchers and decision makers to monitor hundreds of subsidiary variables during the decision-making process.

These reasons justify the current trend in applied modeling with large datasets. The modern econometrician has tools adequate enough to successfully extract information from hundreds of predictor variables and compute more accurate forecasts than ever before. It is noteworthy that these tools mainly do not rely on economic theory in an explicit way; rather they are statistical and consequently atheoretical methods that are used to cover the unfortunate gap between theoretical models and their empirical validation. Within the sum of all possible options, two methods in particular have recently gained ground: dimension reduction and model averaging. Among many others, Bernanke et al (2005), Favero et al (2005), Giannone et al (2004), Stock and Watson (2002, 2005a, 2005b) and Koop and Potter (2004) show how forecasts can be improved over univariate or multivariate autoregressions, using either dynamic factors or Bayesian model averaging (BMA), or both techniques, when a rich dataset is in hand.

In this paper I examine empirically the merit of using factors extracted from a large set of explanatory variables and at the same time implementing Bayesian model averaging/selection in the context of macroeconomic vector autoregressions (VARs). While factor methods have already been examined thoroughly in multivariate models, the challenging task of model averaging/selection is implemented with a stochastic search variable selection algorithm (henceforth SSVS) proposed by George and McCulloch (1993, 1997) and George et al (2008).

The proposed approach is flexible as its output can easily be used for selection of a single best model or model averaging. The SSVS adds to a recent and expanding liter-

ature on different approaches to BMA in VARs (Strachan & van Dijk, 2007; Andersson & Karlsson, 2008). The innovation of the specific prior formulation is that it is more appropriate for VAR models compared to previous model selection priors used in multivariate regressions (Brown et al., 1998, 2002). That is because each right-hand side variable is allowed to enter in all, some, or none of the VAR equations, and not only in all or none of them. The additional advantages come from the fact that this class of restriction search algorithms is extremely simple to use and automated. Furthermore, certain versions of these algorithms can incorporate variable selection when the number of predictors is larger than the number of time series observations.

The following section defines the Bayesian VAR model when many variables are available. Within this "large model approach" the large number of variables is replaced with a small number of factors and several aspects of this approach are discussed. In Section 3, the stochastic restriction search is introduced as a means of efficiently selecting a subset of macroeconomic variables or factors that should be restricted from the VAR specification, based only on the information in the data. The prior specification necessary for model selection is analyzed, as well as the interpretation of model selection probabilities as a special case of BMA. Section 4 outlines the setting of the empirical section (data, forecasting models, prior hyperparameters, and comparison statistics), and the results of the forecasting performance of various VAR specifications. Section 5 concludes the paper with a summary and thoughts for further extension of the basic framework presented in this paper.

2. METHODOLOGY

Let y_t be an $m \times 1$ vector of variables of interest (that we want to forecast) observed for $t = 1, ..., T$. Unlike previous univariate studies (Stock & Watson, 2002, Koop and Potter, 2004), $m > 1$ and I define a forecasting model for y using a general VAR representation

$$y'_{t+1} = \sum_{i=0}^{p_1} y'_{t-i} a_i + w'_t c_0 + \varepsilon'_{t+1} \tag{1}$$

where the parameter matrices a_i and c_0 are of dimensions $m \times m$ and $N \times m$ respectively, y_{t-i}, $i = 1, ..., p_1$,are lagged values of the dependent variable, w_t is a $N \times 1$ vector containing current and lagged values of some exogenous predictor variables, and the

errors are iid Gaussian, $\varepsilon_t \sim N(0, \Sigma)$.This model can be estimated both by OLS and Bayesian methods, provided that the total number of explanatory variables will not exceed the total number of time series observations T. I propose to adopt a Bayesian setting which allows for a unified treatment of this model in high dimensions. For a review of the VAR under standard prior specifications and different sampling methods, the reader is referred to Kadiyala and Karlsson (1997).

Assume we have available observations $x_t = (x_{1t}, ..., x_{nt})'$ on some macroeconomic quantities, where n is large (in the order of hundreds). A popular and simple method to incorporate into an econometric model all the information inherent in a large set of variables, is to reduce their dimension into a lower-dimensional vector of $k \ll n$ latent factors and insert these in the VAR model as explanatory variables

$$x_t' = \lambda f_t' + u_t' \tag{2}$$

$$y_{t+1}' = \sum_{i=0}^{p_1} y_{t-i}' a_i + \sum_{j=0}^{p_2} f_{t-j}' b_j + \varepsilon_{t+1} \tag{3}$$

where f_t is an $k \times 1$ vector of unobserved factors, λ is the matrix of factor loadings and u_t are iid normal errors, $u_t \sim N(0, W)$. In equation (3) the same assumptions hold as in the base model in (1), with the only difference that now $w_t = (f_{t-1}, ..., f_{t-p_2})'$ and $N = k \times (p_2 + 1)$, and the b_j are of appropriate dimensions. For simplicity x_t is demeaned which is equivalent to imposing a constant term in the factor equation, equal to the sample mean $\overline{x} = \frac{1}{T} \sum x_t$ (which in this model coincides both with the MLE of the constant or the mode of its posterior under a diffuse prior). The factors are unobserved quantities and usually it is assumed that they follow a normal distribution with diagonal covariance matrix. One more convention in the factor model literature is to impose the covariance matrix of the innovations, W, to be also diagonal so that (2) reduces to independent equations. Estimation methods vary from principal component analysis (PCA) to full likelihood-based approaches. The ultimate goal of using the factor model is to obtain the factor scores f_t as a valid reduced representation of the manifest vector x_t, so that factor identifiability issues play no actual role here and will not be further discussed.

In terms of the general forecasting VAR model in equation (1), I replace the predictors w_t with the principal components (PC) estimates of the factors $\widehat{F}_t = \left[\widehat{f}_t, \widehat{f}_{t-1}, ..., \widehat{f}_{t-p_2}\right]$, i.e., as if they were observed data. Note that this specification is slightly different from

the dynamic factor model (or factor-augmented VAR) used in Bernanke et al (2005). From their point of view, the dynamic factor model (DFM) is treated as a state-space model, which has the advantage of a probably more efficient one-step estimation of the factors (i.e., along with the parameters of the model) through the Kalman filter algorithm. But this comes at a huge computational cost which makes the application of this model prohibitive in the recursive forecasting setting adopted in this study. After all, Stock and Watson (2005a) have already implemented a large-scale forecasting exercise involving DFMs where they compare several frequentist, full Bayes, and empirical Bayes approaches.

The factors replace the original variables in order to allow richer dynamics and subsequently are allowed to have up to $p_2 + 1$ lags. If the original observed series $x_t = (x_{1t}, ..., x_{nt})'$ were included as predictors then – for a typical macroeconomic dataset with monthly observations on many variables – a degrees of freedom problem would occur if more than one or two lags were assumed. However, even in the case of reducing the dimension of our data with factors the fact that we would ideally allow for many lags does not resolve the problem of overparameterization. For $N = k \times (p_2 + 1)$ larger than 20 the number of all possible models will tend virtually to infinity so that pairwise comparison is practically infeasible using an AIC/BIC-type criterion or prior predictive (marginal) densities and Bayes factors. A reasonable proposed solution from a Bayesian point of view is to use shrinkage subjective priors. For example, the Minnesota prior imposes restrictions on parameters which correspond to higher order lags of y, whereas the prior weight (i.e., the prior mean) for the parameter on the first own lag in each of the m equations is equal to one, and zero on the first lag of the rest $m - 1$ dependent variables. While this approach will work well in VARs which include only lags of the dependent variables, it is difficult to adopt this approach in the models examined here. This happens because there is no theoretical or empirical justification for constructing a subjective prior on exogenous predictor variables, especially if these exogenous variables are latent (constructed) factors.

Introducing any kind of subjective prior information in this model is not an easy task, anyway. These priors may not be specified concretely because of the lack of prior information regarding joint distributions or the large amount of models involved in the analysis. In that respect, subjective prior beliefs require a huge amount of input from the researcher. It is unrealistic to assume that uncertainty about the true model specification can be described meaningfully using ones' own beliefs; hence prior elicitation

should be based mainly on economic theory. The problem with this approach is that in many cases economic theory has empirically proven to be bad guidance in proposing relevant predictors. Stock and Watson (2003) argue that this is the case when forecasting inflation: "the literature does suggest [. . .] variables with the clearest theoretical justification for use as predictors often have scant empirical predictive content."

The discussion so far has focused on the "large-n" case, avoiding to mention anything about how small or large the dimension m of the dependent variable y should be. Although macroeconomic VARs typically contain as dependent variables three or four fundamental quantities that describe the economy, when forecasting, the actual number of variables of interest can grow large. A decision maker would be interested to forecast future values of many series, like production, employment/unemployment, short- and long-term interest rates, consumer and producer price inflation, exchange rates, and many other nominal or real quantities. This is easily handled with the model selection algorithm which is the focus of the next section. The methods described below apply to large VARs in a general sense, that is (i) when the number of predictors n grows large and the number of dependent variables m is small, (ii) when m grows large and n is small, or (iii) when both $m, n \to \infty$, although the empirical application is centered upon the first case.

3. BAYESIAN MODEL SELECTION AND AVERAGING

As was mentioned in the introductory section, when the number of candidate models is too large to enumerate, posterior sampling methods are necessary for the computation of marginal likelihoods for model comparison. Stochastic search algorithms that base on a Markov chain on model space identify regions of high posterior probability and can be used for model selection or to obtain posterior weighted estimates for model averaging. When applied to small models, these algorithms have the ability to search the entire model space, while in large settings only more plausible models are visited. An indicator (zero/one) variable γ, epitomizes the core of Bayesian model selection using stochastic search techniques. Let us define the vector $\gamma = (\gamma_1, ..., \gamma_s)$ as the complete set of indicators, where s is the maximum number of parameters in the model. Then we can proceed by defining a prior $p(\gamma)$ which combined with the likelihood $p(data|\gamma)$, will give zero or one value for each γ_i, $i = 1, ..., s$, from the (updated based on data) posterior distribution $p(\gamma|data)$. This posterior distribution entails all the necessary in-

formation for model selection and averaging. The main idea is to impose the vector of parameters, say $\theta = (\theta_1, ..., \theta_s)$, to have a structure conditional on the values of γ, so that when $\gamma_i = 1$ the associated parameter θ_i will be estimated according to its unrestricted posterior density, and when $\gamma_i = 0$ this would imply $\theta_i = 0$.

There are many ways to implement this general strategy and many alternative methods exist which involve several prior specifications. An analytical review of model averaging and selection is offered in Hoeting et. al (1998) and Chipman et al. (2001). A computationally fast restriction search is described in this section which is based on the SSVS algorithm of George and McCulloch (1993, 1997).

Define $z_t = \left[y_t', y_{t-1}', ..., y_{t-p_1}', w_t'\right]'$, then the VAR model in familiar matrix form is obtained by stacking the row vectors y_{t+1}, z_t and ε_t for $t = 1, ..., T$

$$y = z\phi + \varepsilon, \quad \varepsilon \sim N(0, \Sigma) \tag{4}$$

where $y = \left[y_2', ..., y_{T+1}'\right]'$, $z = [z_1', ..., z_T']'$, $\phi = [a_0, ..., a_{p_1}, c_0]$, and $\varepsilon = \left[\varepsilon_2', ..., \varepsilon_{T+1}'\right]'$. Note that when forecasts are projected h-steps ahead, y is the matrix $y = \left[y_{1+h}', ..., y_{T+h}'\right]'$ (see next section for a definition). Let $n_u = m \times (m \times (p_1 + 1) + k \times (p_2 + 1))$ be the total number of elements in $\varphi = vec(\phi)$. From these elements the m in total constants are always included in the models and admit a typical normal prior of the form

$$(\varphi^c) \sim N\left(\underline{\varphi}^c, vI_m\right) \tag{5}$$

where φ^c is the block of φ which contains the constant terms. Let φ^k be the vector of the remaining $n_\varphi = n_u - m$ parameters in φ which are subject to restriction search and let $\gamma = \left(\gamma_1, ..., \gamma_{n_\phi}\right)$ be the vector of indicator variables associated with the elements of φ^k. Then each element φ_i^k conditional on γ_i, $i = 1, ..., n_\varphi$, follows a scale mixture of normals prior of the form

$$\left(\varphi_i^k | \gamma_i\right) \sim (1 - \gamma_i) N\left(0, \tau_{0i}^2\right) + \gamma_i N\left(0, \tau_{1i}^2\right) \tag{6}$$

The hyperparameters τ_{0i}, τ_{1i} are selected in such a way so that τ_{0i}^2 is small (or even zero) and τ_{1i}^2 is large. Subsequently each parameter φ_i^k is restricted with zero prior mean and very small (or zero) prior variance when $\gamma_i = 0$, while for $\gamma_i = 1$ has a large (locally uninformative) prior variance and in that respect is left unrestricted.

It would not make sense to define the γ_i's if these were defined subjectively and not

updated by the information in the data. Hence a Bernoulli prior on these variables is placed, which updated by the likelihood will result in a conditional posterior which is also Bernoulli. The elements of the vector γ follow an independent Bernoulli $p_i \in (0,1)$ prior of the form

$$(\gamma) \sim \prod p_i^{\gamma_i} (1-p_i)^{(1-\gamma_i)}, \ i = 1, ..., n_\varphi \tag{7}$$

This prior choice reduces computational costs and leads to a posterior density which is easy to derive. In this case $p(\gamma_i = 1) = p_i = 1 - p(\gamma_i = 0)$ so that p_i reflects the prior belief that φ_i^k is large enough and should be left unrestricted. By selecting $p_i < 1/2$, models with an unreasonably large number of parameters are downweighted in order to highlight the significance of parsimonious models. The special case where $p_i = 1/2 \ \forall \ i$, is equivalent to a constant uniform prior $p(\gamma) \equiv 1/2^{n_\varphi}$. This prior is uninformative in the sense that it favors each parameter equally; see Section 4.2 in this paper for more details, and the discussion in Chipman et al. (2001).

The hierarchical mixture prior described above is straightforward to interpret and can be applied virtually to any model for which a normal prior can be specified3 as the conjugate prior that leads to easy derivation of the underlying posterior. A different version of the SSVS is used in Brown et al. (1998) for a multivariate regression model used to predict three variables using 160 predictors. Following the suggestions of George and McCulloch (1997) and Smith and Kohn (1996) they set in equation (6) $\tau_{0i}^2 = 0$ and $\tau_{1i}^2 = g \times \left(z_\gamma' z_\gamma\right)^{-1}$. This prior implies that the first component of the mixture is a Dirac delta function at zero, i.e., a function that puts point mass at zero and hence whenever $\gamma_i = 0$, φ_i^k will be exactly zero. The second component is Zellner's g-prior specification and suggestions for setting uninformative values of g (although in a univariate context) are given in Fernandez, Ley, and Steel (2001). An updated and computationally more efficient version of this prior specification appears in Brown et al. (2002), where more variables than observations can be handled. The shortcoming of their approach is that it is able to treat each equation in the VAR individually, but instead is choosing the variables in z which are more probable to be included in *all* VAR equations together. Put simply, if, say, z contains only the first lag of the dependent variables, then the latter approach will allow the y_{it-1} to be a predictor of the whole vector y_t, while the approach proposed here $y_{it-1}$to explain the dependent variable in equation j of the VAR (denoted y_{jt}), but not the dependent variable in the l-th VAR equation (denoted y_{lt}). Nevertheless, the Brown et al. (2002) implementation of the SSVS algorithm is a valuable complement to the one used here, and undoubtedly a useful tool in

empirical analysis with focus on prediction.

Smith and Kohn (2002) extend the stochastic search for parameter restrictions to the covariance matrix of longitudinal data. George et al. (2008) apply their idea to the covariance matrix of structural VARs: motivated by the fact that identifying restrictions on the covariance are usually imposed on the elements of a reparametrization of Σ, they focus on restricting the elements of the $m \times m$ upper triangular matrix Ψ satisfying

$$\Sigma^{-1} = \Psi'\Psi \tag{8}$$

They then derive a mixture of normals prior, as in equation (6), for the nondiagonal elements of Ψ, while the diagonal is integrated out with a gamma prior. Matrix Ψ has the form

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1m} \\ 0 & \psi_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \psi_{mm} \end{bmatrix} \tag{9}$$

so let $\boldsymbol{\psi} = (\psi_{11}, ..., \psi_{mm})'$ and $\eta = (\eta_2', ..., \eta_m')' = \left(\psi_{12}, \psi_{13}, \psi_{23}, ..., \psi_{(m-1)m}\right)'$be the vectors of the diagonal and upper diagonal elements respectively, where $\eta_j = \left(\psi_{1j}, ..., \psi_{(j-1)j}\right)'$ for $j = 2, ..., m$. Let $\omega_j = \left(\omega_{1j}, ..., \omega_{(j-1)j}\right)'$ be a vector of 0-1 variables so that each element of η_j has prior conditional on ω_j of the form

$$\left(\eta_{ij}|\omega_{ij}\right) \sim (1 - \omega_{ij})\, N\left(0, \kappa_{0ij}^2\right) + \omega_{ij} N\left(0, \kappa_{1ij}^2\right) \tag{10}$$

for $i = 1, ..., j-1$ and $j = 2, ..., m$. As in the case of the vector γ, assume that the elements of the vector $\omega = (\omega_2', ..., \omega_m')'$ are independent Bernoulli $q_{ij} \in (0, 1)$ random variables so that

$$(\omega) \sim \prod\nolimits_i \prod\nolimits_j q_{ij}^{\omega_{ij}} \left(1 - q_{ij}\right)^{(1-\omega_{ij})} \tag{11}$$

For $i = 2, ..., m$, each ψ_{ii} has a gamma prior density

$$\left(\psi_{ii}^2\right) \sim Gamma\left(\alpha_i, \beta_i\right)$$

For more information on these priors the reader is referred to the analytical calculations of George et al. (2008) where it is shown that finding restrictions on the covariance matrix based solely on the data provides an attractive alternative to identifying

restrictions imposed in structural VARs. It should be clear from the prior specification that the SSVS is an intuitive extension of the Bayesian conjugate (normal – inverse Wishart) prior. In the empirical application I adopt a fast sampling scheme (see Section 4.2) to draw from the posteriors of c and x, which makes computation feasible in multivariate models. The parameter posteriors are given in detail in Appendix A (Technical Appendix). Although selection of prior hyperparameters seems to be fairly automatic in this setting, prior elicitation is an important factor in model selection.

4. EMPIRICAL APPLICATION

4.1 Data

I use the Stock and Watson (2005b) dataset which is an updated version of the Stock and Watson (2002) dataset that is widely used in empirical applications. This version consists of 132 monthly variables pertaining to the US economy measured from 1960:01 to 2003:12. The 132 predictors can be grouped in 14 categories: real output and income; employment and hours; real retail, manufacturing, and trade sales; consumption; housing starts and sales; real inventories; orders; stock prices; exchange rates; interest rates and spreads; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous. The data were transformed to eliminate trends and nonstationarities. All the data and transformations are summarized in Appendix B.

4.2 Selection of prior hyperparameters

Implementation of Bayesian model selection requires all the priors to be proper, as the ones described in Section 3. Noninformative improper priors are not suitable to calculate Bayes factors and posterior model probabilities. Even though there are certain methods which overcome this difficulty (BIC approximations, intrinsic, or fractional Bayes factors), the standard practice in the Bayesian model selection literature is to use only proper priors. This does not necessarily mean that noninformative proper priors cannot be specified. It is easy to choose the hyperparameters in such a way that all the priors are locally noninformative.

Selection of τ_{0i}, τ_{1i} and κ_{0ij}, κ_{1ij} can be made along the guidelines of Chipman et al. (2001, p. 86). For instance, given a non-negative scalar threshold ξ_i, higher posterior

weighting can be allocated to those values of γ for which $\left|\varphi_i^k\right| > \xi_i$ when $\gamma_i = 1$, iff τ_{0i}, τ_{1i} satisfy

$$\log\left(\frac{\tau_{1i}/\tau_{0i}}{\tau_{0i}^{-1} - \tau_{1i}^{-1}}\right) = \xi_i^2$$

A similar argument can be made for the choice of κ_{0ij} and κ_{1ij}. Alternatives for a more objective selection of these hyperparameters exist, but at the cost of a substantial increase in computational calculations. The first one is to use empirical Bayes criteria in the spirit of George and Foster (2000), while a fully Bayes approach would require to place an inverted-Gamma hyperprior on each τ_{0i}, τ_{1i} and κ_{0ij}, κ_{1ij}. Selection based on the formula above is a simple task which can easily be implemented in large models. George et al. (2008) argue that even if the restriction search algorithm is not effective in selecting the correct restrictions on ϕ, the results can still be used to obtain improved forecasts.

The only source of difficulty may arise in eliciting the hyperparameters of the Bernoulli random variables γ (similarly ω). The prior structure that appears in equation (7) (similarly in equation (11)) is an "independence prior," in the sense that each element of γ (ω) is assumed to be independent of the rest. This simplification makes it difficult to account for similarities or differences between models when the correlation between the explanatory variables is high. While priors that "dilute" probability across neighborhoods of similar models (Chipman et al., 2001; Yuan & Lin, 2005) are able to correct this shortcoming, it is preferable to use an orthogonal transformation of the variables in z, by applying a singular value decomposition. This allows exploring the model space in considerably less iterations, which subsequently decreases the computational cost in multivariate models. Hence, in the forecasting exercise, I apply the restriction search to the model

$$y_{T+h} = G_T \mu_h + \varepsilon_{T+h}$$

where $G = zH$ are orthogonal variables and $\mu = H^{-1}\phi$; see Koop and Potter (2004). This approach will speed up computations, even though orthogonality does not lead to posterior independence of elements of γ. The default choice $p_i = 1/2$ in equation (7) and $q_{ij} = 1/2$ in equation (11) may result in a uniform prior, but this would not be a noninformative prior about model size. A rule of thumb is that if the researcher anticipates many (few) restrictions on the model then the choice should be $p_i, q_{ij} < 1/2$ ($p_i, q_{ij} > 1/2$). Prior sensitivity analysis using real and simulated data showed that $p_i = q_{ij} = 1/2$ is able to identify restrictions quite well and hence is left as the default

reasonable choice.

Following the suggestions of George et al. (2008) and George and McCulloch (1997), I adopt a fast sampling scheme for γ and ω, which requires to set τ_{0i} and κ_{0ij} small, but different from 0. According to the preceding discussion in this subsection and the absence of prior beliefs about specific parameters I set $\tau_{0i} = \tau_0 = 0.01$, $\tau_{1i} = \tau_1 = 70$ for all $i = 1, ..., n_\varphi$, and $\kappa_{0ij} = \kappa_0 = 0.01$, $\kappa_{1ij} = \kappa_1 = 30$ for all $j = 2, ..., m$ and $i = 1, ..., j-1$. For the intercept term, the typical normal prior has mean $\underline{\varphi}^c = 1$ and variance $v = 100$. A default noninformative choice for the parameters of the Gamma density is $\alpha_i, \beta_i = 0.01$.

4.3 Implementation of Bayesian Model Averaging/Selection

At this point, as it is practically impossible to summarize model selection results from the recursive forecasting exercise, I summarize the average posterior probability of some of the variables in the dataset without extracting factors, i.e., replacing w_t with $x_t = (x_{1t}, ..., x_{nt})'$ in specification (1), and using the full sample of observations from 1960:1 to 2003:12. I consider a New Keynesian VAR with three variables (unemployment, consumer price index, and federal funds rate) regressed on an intercept, 14 autoregressive lags, and the remaining 129 variables in the dataset which are used as exogenous predictors. This gives a total of $129 + 13 \times 3 = 168$ right-hand side variables (excluding the intercept which is always included) to choose from in each equation. The horizon chosen in this illustration is $h = 12$. The unemployment and interest rate are transformed to stationarity by taking first differences. The consumer price index is transformed by taking the second difference of the logarithm.

A parameter should either be included or excluded, hence the number of all possible models is 2^{168} in each VAR equation and $2^{168\times 3} = 5.2e + 151$ in total. The BMA posterior probabilities are computed for each parameter $i = 1, ..., n_\varphi$ as

$$E(\gamma_i|y) = \frac{1}{S}\sum_{s=1}^{S}\gamma_i^{(s)}$$

where S is the total number of iterations from the posterior sampler, and $\gamma_i^{(s)}$ are draws from the conditional posterior of γ_i. This suggests that the average probability is actually the proportion of models visited by the Gibbs sampler, which contain the corresponding variable. Exactly similar inference and interpretation holds for the parameters

ω, although these index elements of the covariance matrix and not columns of predictors in mean VAR equation.

Tables C1 and C2 summarize the results for those predictor variables and own lags, respectively, that have the highest probabilities. Variables which had average posterior probability less than 0.5 in all of the three equations are not included at all in the tables. Each element in these tables is the BMA posterior probability and can be interpreted simply as the probability that the corresponding right-hand side variable should be included. For this specific application the variables are not orthogonalized in order to retain the interpretation of the probabilities as the amount of belief that the respective variable is included in the model. The results are based on 150,000 iterations with a burn-in period of 50,000, which leaves 100,000 draws to evaluate the posterior of γ. Elicitation of prior hyperparameters is based on the values described in Section 4.2.

Note that the probabilities ω for Ψ are 0.52, 1, and 1 for each of the upper diagonal elements ψ_{12}, ψ_{13}, and ψ_{23} respectively. Once all these probabilities are available, it is straightforward to interpret them. This output can be used to implement BMA if all variables contribute to the final forecast according to their probability, no matter how high or low this probability is. Looking for example at Table C1, the spread of the 10-year interest rate from the federal funds rate variable will contribute to the final forecast of the unemployment rate, the consumer price index, and the interest rate in 100, 86.1, and 100%of the occasions (models visited by the sampler), respectively. In contrast the same output can be used to select the best single model. Barbieri and Berger (2004) show that in the context of Bayesian model selection the optimal model is the median probability model. According to this result, only the variables which have average probability larger than 0.5 in each equation will be unrestricted. These probabilities are presented in Tables C1 and C2. Hence, in this "best" model, the 1, 5, and 10-year interest rate spreads should be included in all three equations, while capacity utilization should enter only the unemployment equation.

The results presented in Table C1 are also subject to economic interpretation. Space restrictions, however, do not allow further analysis in this study. Structural interpretation is not the main focus, but forecast improvement is. This is an issue examined in the following section.

4.4 Forecasting in Large VAR Models

The first estimation period is set to 1960:1 and a simulated real-time forecasting of y_{t+h} is done from 1983:1 through 2003:12-h, for horizons $h = 1, 6,$ and 12. Each VAR model has eight dependent variables of interest (with their short mnemonic from the dataset in parentheses): Personal Income ($A0M052$), Industrial Production ($IPS10$), Employment Rate ($CES002$), Unemployment Rate ($LHUR$), 3-month Treasury Bill Rate ($FYGM3$), Producer Price Index ($PWFSA$), Consumer Price Index ($PUNEW$), and PCE Deflator ($GMDC$). This leaves a total of 124 variables to explore their predictive content. As mentioned earlier, all the variables are transformed to stationarity, a fact that implies a specific transformation of the variable y_{t+h} proper for forecasting. Let v_{it} denote the untransformed value of y_{it} for each of the eight monthly dependent variables i, then $y_{it+h} = (1200/h)\log(v_{it+h}/v_{it})$ for $i = (A0M052, IPS10, CES002)$, $y_{it+h} = v_{it+h} - v_{it}$ for $i = (LHUR, FYGM3)$, and $y_{it+h} = (1200/h)\{\log(v_{it+h}/v_{it}) - h\Delta\log(v_{it})\}$ for $i = (PWFSA, PUNEW, GMDC)$.

The principal components are estimated from the 124 variables in the dataset using the same sample period as the VAR. Several multivariate forecasting exercises in the literature (cf. Stock & Watson, 2002) focus on finding the best performing model. In contrast, here the main challenge is to improve forecasts when the number of predictors grows large and the researcher has no prior information about which is the correct model size. Thus, the maximum potential number of factors and lags is deliberately set to large, "uninformative" values. In particular, 10 principal components ($k = 10$) are extracted from the factor model in equation (2), while the VAR specification in equation (3) contains an intercept, 13 autoregressive lags ($p_1 = 12$), and 13 lagged factors ($p_2 = 12$). This gives a maximum of 221 (plus the intercepts, which are unrestricted) potential predictors of each of the 8 dependent variables. For the purpose of the empirical application forecasts are computed from: (i) VAR with SSVS and model averaging, (ii) VAR with SSVS and model selection, and (iii) VAR estimated using OLS with selection of predictors with the Bayesian information criterion (which has a larger penalty for less parsimonious models relative to the Akaike information criterion, and is a rough approximation to the Bayes factors). The predictors in the latter method are orthogonalized and the total number of possible models considered is equal to the maximum number of right-hand side variables and subsequently selection of the best model is implemented in a finite number of calculations.

An appropriate common way to quantify out-of-sample forecasting performance is to compute the root mean square forecast error (RMSFE) statistic for each forecast horizon h:

$$RMSFE_{ij}^{h} = \sqrt{\sum_{t=1982:12}^{2003:12-h} \left(y_{i,t+h}^{*} - \widetilde{y}_{i,t+h,j}\right)^{2}} \tag{12}$$

where $y_{i,t+h}^{*}$ is the realized (observed) value of y at time $t+h$ for the i-th series, and $\widetilde{y}_{i,t+h,j}$ is the mean of the posterior predictive density at time $t+h$, for the i-th series, from the j-th forecasting model. The RMSFE of each model is reported relative to the RMSFE of a benchmark VAR with an intercept and seven lags of the dependent variables, estimated with OLS

$$rRMSFE_{ij}^{h} = \frac{RMSFE_{ij}^{h}}{RMSFE_{iVAR(7)}^{h}} \tag{13}$$

This VAR(7) model is not chosen because of its higher forecasting ability compared to other alternatives. Following the standard convention in the literature an AR(2) model would be a better candidate to serve as the benchmark model. But note that the VAR(7) is nested to the VAR with factors, which will give a better picture of whether the restrictions found by the SSVS are actually the ones that will lead to reduced RMSFE statistics, compared to a more parsimonious alternative. The forecasting performance of the models based on the relative RMSFE for horizons $h = 1, 6, 12$, is summarized in Table C3. These are the averaged values of the RMSFEs over the forecast period, 1983:1 through 2003:12-h.

The results are encouraging about the application of the restriction search algorithm in large models. In most occasions the BMA and Bayesian model selection give improved results compared to the BIC selection. Note that the improvement is not only due to the fact that the models of interest contain more predictors than the benchmark model. It is noteworthy that in some occasions only lags of the dependent variable are selected from the restriction search, while for most samples the number of important lagged factors, for each dependent variable, is not more than five. This is supported by the fact that the average RMSFE (results not reported here) of the large VAR with factors but without selection of predictors (i.e., a heavily overparametrized model) is, as expected, extremely high relative to the VAR(7). An important feature of the restriction search algorithm applied to the specific VAR is that the forecasts from Bayesian model selection are better than the forecasts from BMA. The practical difference of the two approaches is that BMA shrinks the posterior means of the parameter with low probability toward

zero, while Bayesian model selection imposes that these parameters (with probability less than 0.5) will be exactly zero.

5. CONCLUSIONS

This paper addresses the forecasting performance of Bayesian VAR models with many predictors using a flexible prior structure which leads to output that can be used for model selection and model averaging. For eight U.S. monthly macroeconomic variables of interest forecasting accuracy is improved over least squares estimation and selection of predictors using the Bayesian information criterion. Without arguing that the choice of prior hyperparameters was the best possible and done with a strict "objective" criterion (like in other BMA applications, see Fernandez et al., 2001), the gains from the standard automated choices are appreciable. As already mentioned, there are many proposals in the Bayesian literature for efficient elicitation of prior hyperparameters for model selection and some of them were discussed in the paper. Nevertheless, the merit of the SSVS for VAR models lies in its simplicity and intuitive interpretation.

With regard to other macroeconometric specifications, the flexibility of the restriction search algorithm suggests many interesting extensions. Firstly, note that it is straightforward to adopt it in general piecewise-linear multivariate regressions that allow for thresholds, Markov switching or structural breaks; an interesting area for future research. Secondly, I only considered the case where the number of dependent variables, m, is small and the number of predictors grows large. But as already mentioned the restriction search algorithm may also be used when the number of dependent variables grows large. Banbura, Giannone, and Reichlin (2007) examine this case using shrinkage priors and find huge gains from this large VAR specification. Lastly, an interesting direction for future research would be the empirical application of the restriction search algorithm in the Bayesian dynamic factor model. This approach will probably improve forecasting performance and impulse response analysis in DFMs that lack parsimony (cf. Bernanke et al., 2005 and Stock & Watson, 2005b).

ACKNOWLEDGMENTS

The author would like to thank Gary Koop and participants and discussants at several workshops and conferences for helpful comments. Part of this research was supported by the Department of Economics, University of Strathclyde, which I gratefully acknowledge.

REFERENCES

Andersson, M. K., & Karlsson, S. (2008). Bayesian forecast combination for VAR models. In: S. Chib, W. Griffiths, G. Koop & D. Terrell (Eds), *Bayesian Econometrics* (pp. 501–524). Advances in Econometrics, Vol. 23. Oxford: Elsevier.

Banbura, M., Giannone, D., & Reichlin, L. (2007). Bayesian VARs with large panels. London: CEPR.

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32, 870–897.

Bernanke, B. S., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50, 525–546.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of*

Economics, 120, 387–422.

Brown, P. J., Vannucci, M., & Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60, 627–641.

Brown, P. J., Vannucci, M., & Fearn, T. (2002). Bayes model averaging with selection of regressors. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64, 519–536.

Chipman, H., George, E. I., & McCulloch, R. E. (2001). The practical implementation of Bayesian model selection. In: P. Lahiri (Ed.), *Model selection* (pp. 67–116).

Lecture Notes –Monograph Series, Vol. 38. Beachwood, OH: Institute of Mathematical Statistics.

Favero, C. A., Marcellino, M., & Neglia, F. (2005). Principal components at work: The empirical analysis of monetary policy with large datasets. *Journal of Applied*

Econometrics, 20, 603–620.

Fernandez, C., Ley, E., & Steel, M. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100, 381–427.

George, E. I., & Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, 87, 731–747.

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

George, E. I., & McCulloch, R. E. (1997). Approaches to Bayesian variable selection. *Statistica Sinica*, 7, 339–379.

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

Gianone, D., Reichlin, L., & Sala, L. (2004). Monetary policy in real time (and comments). In: M. Gertler & K. Rogoff (Eds), *NBER Macroeconomics Annual 2004* (pp. 161–224). Cambridge: The MIT Press.

Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1998). Bayesian model averaging: A tutorial. *Statistical Science*, 14, 382–417.

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

Koop, G., & Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal*, 7, 550–565.

Smith, M., & Kohn, R. (1996). Nonparametric regression using Bayesian variable section. *Journal of Econometrics*, 75, 317–343.

Smith, M., & Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97, 1141–1153.

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

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41, 788–829.

Stock, J. H., & Watson, M. W. (2005a). Forecasting with many predictors. Unpublished Manuscript. Princeton University, Princeton, NJ (prepared for The Handbook of Economic Forecasting).

Stock, J. H., & Watson, M. W. (2005b). Implications of factor models for VAR analysis. Unpublished Manuscript. Princeton University, Princeton, NJ.

Strachan, R. W., & van Dijk, H. K. (2007). Bayesian model averaging in vector autoregressive processes with an investigation of stability of the US great ratios and risk of a liquidity trap in the USA, UK and Japan. *Econometric Institute Report* EI 2007-11. Erasmus University Rotterdam, the Netherlands.

Yuan, M., & Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100, 1215–1225.

APPENDICES

A TECHNICAL APPENDIX – A GIBBS SAMPLER FOR SSVS IN VAR MODELS

The priors described in Section 3 combined with the likelihood function of a VAR model, will allow us to derive and draw from the full conditional distributions. The likelihood of the VAR model $y = z\phi + \varepsilon$, $\varepsilon \sim N(0, \Sigma)$ with $\Sigma^{-1} = \Psi'\Psi$, is

$$\begin{aligned} L(y|\phi, \Psi) &\propto |\Psi|^{-T} \exp\left\{-\frac{1}{2} tr\left[\Psi'(y - z\phi)'(y - z\phi)\Psi\right]\right\} \\ &= |\Psi|^{-T} \exp\left\{-\frac{1}{2}\left(\phi - \widehat{\phi}\right)'\left[\Psi\Psi' \otimes (z'z)\right]\left(\phi - \widehat{\phi}\right)\right. \\ &\quad \left. -\frac{1}{2} tr\left[\left(y - z\widehat{\phi}\right)'\Psi'\Psi\left(y - z\widehat{\phi}\right)\right]\right\} \end{aligned}$$

where $\widehat{\phi}$ is the MLE of ϕ. This form of the likelihood function allows to derive the posterior of the ϕ parameters. In order to derive the posterior of the elements of Ψ we need to first rewrite the likelihood function in convenient form. Define $S(\phi) = (y - z\phi)'(y - z\phi)$ and write $S(\phi) = s_{ij}$. For $j = 2, ..., m$ define the $(m-1)$ vectors $s_j = \left(s_{1j}, ..., s_{(j-1)j}\right)'$ containing the upper diagonal elements of $S(\phi)$, and the $(m-1)$ matrices S_j containing the upper left $j \times j$ submatrix of $S(\phi)$. Define also $\rho_1 = s_{11}$ and $\rho_i = |S_i| / |S_{i-1}| = s_{ii} - s_i' S_{i-1}^{-1} s_i$ for $i = 2, ..., m$. The likelihood function now cam take the following form

$$L(y|\phi, \Psi) \propto \prod_{i=1}^{m} (\psi_{ii})^T \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m} \psi_{ii}^2 \rho_i + \sum_{j=2}^{m} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)' S_{j-1}\left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)\right]\right\}$$

Now define $D = diag\left\{h_1, ..., h_{n_\varphi}\right\}$ with

$$h_i = \begin{cases} \tau_{0i}, \text{ if } \gamma_i = 0 \\ \tau_{1i}, \text{ if } \gamma_i = 1 \end{cases}, \text{ for } i = 1, ..., n_\varphi$$

and, similarly, define $D_j = diag\left\{h_{1j}, ..., h_{(j-1)j}\right\}$ with

$$h_{ij} = \begin{cases} \kappa_{0ij}, \text{ if } \omega_{ij} = 0 \\ \kappa_{1ij}, \text{ if } \omega_{ij} = 1 \end{cases}$$

for $i = 1, ..., j$ and $j = 2, ..., m$. Then we can rewrite equations (6) and (10) in the main text, as

$$\begin{aligned} \left(\varphi_i^k|\gamma\right) &\sim N\left(0, DD\right) \\ \left(\eta_j|\omega_j\right) &\overset{iid}{\sim} N_{j-1}\left(0, D_j D_j\right) \end{aligned}$$

respectively. Denote the combined prior of the unrestricted coefficients φ^c and the restricted coefficients φ^k as $\varphi \sim N\left(\underline{\varphi}, \underline{V}\right)$. Given starting values, model parameters are drawn from their conditionals for $r = 1, ..., R$ iterations:

1. Draw $\left(\psi^{(r)}|\eta^{(r-1)}, \omega^{(r-1)}, \gamma^{(r-1)}, \varphi^{(r-1)}, data\right)$ by sampling each element from the Gamma distribution

$$\psi_{ii}^2 \sim Gamma\left(\alpha_i + \frac{1}{2}T, B_i\right)$$

where

$$B_i = \begin{cases} \beta_1 + \frac{1}{2}s_{11} & \text{for } i = 1 \\ \beta_i + \frac{1}{2}\left[s_{ii} - s_i'\left(S_{i-1} + \left(D_i D_i\right)^{-1}\right)^{-1} s_i\right] & \text{for } i = 2, ..., m \end{cases}$$

2. Draw $\left(\eta^{(r)}|\psi^{(r)}, \gamma^{(r-1)}, \varphi^{(r-1)}, \omega^{(r-1)}, data\right)$ by sampling each element from the Normal distribution

$$\left(\eta_j\right) \sim N_{j-1}\left(\mu_j, \Delta_j\right)$$

where for $j = 2, ..., m$.

$$\begin{aligned} \mu_j &= -\psi_{jj}\left\{S_{j-1} + \left(D_j D_j\right)^{-1}\right\}^{-1} s_j \\ \Delta_j &= \left\{S_{j-1} + \left(D_j D_j\right)^{-1}\right\}^{-1} \end{aligned}$$

3. Draw $\left(\omega^{(r)}|\eta^{(r)}, \psi^{(r)}, \gamma^{(r-1)}, \varphi^{(r-1)}, data\right)$ by sampling each element from the Bernoulli

distribution

$$(\omega_{ij}) \sim Bernoulli\left(1, \frac{u_{1ij}}{u_{1ij}+u_{2ij}}\right)$$

where for $j = 2, ..., m$ and $i = 1, ..., j-1$

$$\begin{aligned} u_{1ij} &= \frac{1}{\kappa_{0ij}} \exp\left(-\frac{\psi_{ij}^2}{2\kappa_{0ij}^2}\right) q_{ij} \\ u_{2ij} &= \frac{1}{\kappa_{1ij}} \exp\left(-\frac{\psi_{ij}^2}{2\kappa_{1ij}^2}\right) (1-q_{ij}) \end{aligned}$$

4. Draw $\left(\varphi^{(r)}|\eta^{(r)}, \psi^{(r)}, \omega^{(r)}, \gamma^{(r-1)}, data\right)$ by sampling $\varphi = vec(\phi)$ from the Normal distribution

$$(\varphi) \sim N_{n_u}(\mu, \Delta)$$

where

$$\begin{aligned} \mu &= \left\{(\Psi\Psi') \otimes (z'z) + \underline{V}^{-1}\right\}^{-1} \left\{((\Psi\Psi') \otimes (z'z))\,\widehat{\varphi} + \underline{V}^{-1}\underline{\varphi}\right\} \\ \Delta &= \left\{(\Psi\Psi') \otimes (z'z) + \underline{V}^{-1}\right\}^{-1} \end{aligned}$$

where $\widehat{\varphi}$ is the vector occuring from stacking the elements of the matrix of MLE coefficients, i.e. $\widehat{\varphi} = vec\left(\widehat{\phi}\right) = vec\left((z'z)^{-1} z'y\right)$.

5. Draw $\left(\gamma^{(r)}|\eta^{(r)}, \psi^{(r)}, \omega^{(r)}, \varphi^{(r)}, data\right)$ by sampling each element from the Bernoulli density

$$(\gamma_i) \sim Bernoulli\left(1, \frac{u_{1i}}{u_{1i}+u_{2i}}\right)$$

1. where for $i = 1, ..., n_u$

$$\begin{aligned} u_{1i} &= \frac{1}{\tau_{0i}} \exp\left(-\frac{\varphi_i^2}{2\tau_{0i}^2}\right) p_i \\ u_{2i} &= \frac{1}{\tau_{1i}} \exp\left(-\frac{\varphi_i^2}{2\tau_{1i}^2}\right) (1-p_i). \end{aligned}$$

B DESCRIPTION OF DATA

This table lists the 132 variables in the dataset used. The third column indexes the respective transformation applied to each of the variables to ensure stationarity (at least approximately). Let vt and xt be the untransformed value and transformed values respectively, then there are five cases:(1) lv: $x_t = v_t$ (level), (2) $\ln$: $x_t = \log(v_t)$ (logarithm), (3) Δlv: $x_t = v_t - v_{t-1}$ (first difference), (4) $\Delta \ln$: $x_t = \log(v_t/v_{t-1})$ (growth rate), and (5) $\Delta^2 \ln$: $x_t = \Delta \log(v_t/v_{t-1})$ .This table is from Stock and Watson (2005b) and the reader should seek in this reference the original source of the data.

#	Mnemonic	Trans	Description
1	A0M052	$\Delta \ln$	Personal income (ar, bil. chain 2000 $)
2	A0M051	$\Delta \ln$	Personal income less transfer payments (ar, bil. chain 2000 $)
3	A0M224	$\Delta \ln$	Real consumption ($A0M224/GMDC$)
4	A0M057	$\Delta \ln$	Manufacturing and trade sales (mil. chain 1996 $)
5	A0M059	$\Delta \ln$	Sales of retail stores (mil. chain 2000 $)
6	IPS10	$\Delta \ln$	Industrial production index - total index
7	IPS11	$\Delta \ln$	Industrial production index - products, total
8	IPS299	$\Delta \ln$	Industrial production index - final products
9	IPS12	$\Delta \ln$	Industrial production index - consumer goods
10	IPS13	$\Delta \ln$	Industrial production index - durable consumer goods
11	IPS18	$\Delta \ln$	Industrial production index - nondurable consumer goods
12	IPS25	$\Delta \ln$	Industrial production index - business equipment
13	IPS32	$\Delta \ln$	Industrial production index - materials
14	IPS34	$\Delta \ln$	Industrial production index - durable goods materials
15	IPS38	$\Delta \ln$	Industrial production index - nondurable goods materials
16	IPS43	$\Delta \ln$	Industrial production index - manufacturing17
17	IPS307	$\Delta \ln$	Industrial production index - residential utilities
18	IPS306	$\Delta \ln$	Industrial production index - fuels
19	PMP	lv	NAPM production index (percent)

#	Mnemonic	Trans	Description
20	A0M082	Δlv	Capacity utilization (mfg)
21	LHEL	Δlv	Index of help-wanted advertising in newspapers (1967=100;sa)
22	LHELX	Δlv	Employment: ratio; help-wanted ads/ no. unemployed clf
23	LHEM	Δlv	Civilian labor force: employed, total (thous.)
24	LHNAG	Δlv	Civilian labor force: employed, nonagricultural industries (thous.)
25	LHUR	Δlv	Unemployment rate: all workers, 16 years & over (%)
26	LHU680	Δlv	Unemployment by duration: average (mean) duration in weeks
27	LHU5	$\Delta \ln$	Unemployment by duration: persons unemployed less than 5 wks (thous.)
28	LHU14	$\Delta \ln$	Unemployment by duration: persons unemployed 5 to 14 wks (thous.)
29	LHU15	$\Delta \ln$	Unemployment by duration: persons unemployed 15 wks + (thous.)
30	LHU26	$\Delta \ln$	Unemployment by duration: persons unemployed 15 to 26 wks (thous.)
31	LHU27	$\Delta \ln$	Unemployment by duration: persons unemployed 27 wks + (thous.)
32	A0M005	$\Delta \ln$	Average weekly initial claims, unemployment insurance (thous.)
33	CES002	$\Delta \ln$	Employees on nonfarm payrolls - total private
34	CES003	$\Delta \ln$	Employees on nonfarm payrolls - goods-producing
35	CES006	$\Delta \ln$	Employees on nonfarm payrolls - mining
36	CES011	$\Delta \ln$	Employees on nonfarm payrolls - construction37
38	CES017	$\Delta \ln$	Employees on nonfarm payrolls - durable goods
39	CES033	$\Delta \ln$	Employees on nonfarm payrolls - nondurable goods
40	CES046	$\Delta \ln$	Employees on nonfarm payrolls - service-providing
41	CES048	$\Delta \ln$	Employees on nonfarm payrolls - trade, transportation, and utilities

#	Mnemonic	Trans	Description
42	CES049	$\Delta \ln$	Employees on nonfarm payrolls - wholesale trade
43	CES053	$\Delta \ln$	Employees on nonfarm payrolls - retail trade
44	CES088	$\Delta \ln$	Employees on nonfarm payrolls - financial activities
45	CES140	$\Delta \ln$	Employees on nonfarm payrolls - government
46	A0M048	$\Delta \ln$	Employee hours in nonagricultural establishments (ar, bil. hours)
47	CES151	lv	Average weekly hours of production or nonsupervisory workers on private nonfarm payrolls
48	CES155	Δlv	Average weekly hours of production or nonsupervisory workers on private nonfarm payrolls
49	AOM001	lv	Average weekly hours: manufacturing (hours)
50	PMEMP	lv	NAPM employment index (percent)
51	HSFR	$\ln$	Housing starts: nonfarm (1947-58); total farm
52	HSNE	$\ln$	Housing starts: Northeast (thousands of units)
53	HSMW	$\ln$	Housing starts: Midwest (thousands of units)
54	HSSOU	$\ln$	Housing starts: South (thousands of units)55
56	HSBR	$\ln$	Housing authorized: total new priv housing units (thousands)
57	HSBNE	$\ln$	Houses authorized by build. permits: Northeast (thousands of units)
58	HSBMW	$\ln$	Houses authorized by build. permits: Midwest (thousands of units)
59	HSBSOU	$\ln$	Houses authorized by build. permits: South (thousands of units)
60	HSBWST	$\ln$	Houses authorized by build. permits: West (thousands of units)
61	PMI	lv	Purchasing managers' index (sa)
62	PMNO	lv	NAPM new orders index (percent)
63	PMDEL	lv	NAPM vendor deliveries index (percent)
64	PMNV	lv	NAPM inventories index (percent)
65	A0M008	$\Delta \ln$	Mfrs' new orders, consumer goods and materials (bil. chain 1982 $)

#	Mnemonic	Trans	Description
66	A0M007	$\Delta \ln$	Mfrs' new orders, durable goods industries (bil. chain 2000 $)
67	A0M027	$\Delta \ln$	Mfrs' new orders, nondefense capital goods (mil. chain 1982 $)
68	A1M092	$\Delta \ln$	Mfrs' unfilled orders, durable goods indus. (bil. chain 2000 $)
69	A0M070	$\Delta \ln$	Manufacturing and trade inventories (bil. chain 2000 $)
70	A0M077	Δlv	Ratio, mfg. and trade inventories to sales (based on chain 2000 $)
71	FM1	$\Delta^2 \ln$	Money stock: M1 (bil$,sa)
72	FM2	$\Delta^2 \ln$	Money stock: M2 (bil$,sa)
73	FM3	$\Delta^2 \ln$	Money stock: M3 (bil$,sa)
74	FM2DQ	$\Delta \ln$	Money supply - M2 in 1996 dollars (bci)
75	FMFBA	$\Delta^2 \ln$	Monetary base, adjusted for reserve requirement changes(mil$,sa)
76	FMRRA	$\Delta^2 \ln$	Depository inst. reserves: total, adjusted for reserve req changes (mil$,sa)
77	FMRNBA	$\Delta^2 \ln$	Depository inst. reserves: non-borrowed, adj reserve req changes (mil$,sa)
78	FCLNQ	$\Delta^2 \ln$	Commercial & industrial loans outstanding in 1996 dollars (bci)
79	FCLBMC	lv	Wkly rp lg com'l banks:net change com'l & indus loans (bil$,saar)
80	CCINRV	$\Delta^2 \ln$	Consumer credit outstanding – non-revolving
81	A0M095	Δlv	Ratio, consumer installment credit to personal income (pct.)
82	FSPCOM	$\Delta \ln$	S&P's common stock price index: composite (1941-43=10)
83	FSPIN	$\Delta \ln$	S&P's common stock price index: industrials (1941-43=10)
84	FSDXP	Δlv	S&P's composite common stock: dividend yield (% per annum)

#	Mnemonic	Trans	Description
85	FSPXE	$\Delta \ln$	S&P's composite common stock: price-earnings ratio (%)
86	FYFF	Δlv	Interest rate: Federal funds (effective) (% per annum)87
88	FYGM3	Δlv	Interest rate: u.s. Treasury bills, sec market, 3-mo.(% per annum)
89	FYGM6	Δlv	Interest rate: u.s. Treasury bills, sec market, 6-mo.(% per annum)
90	FYGT1	Δlv	Interest rate: u.s. Treasury const maturities, 1-yr.(% per annum)
91	FYGT5	Δlv	Interest rate: u.s. Treasury const maturities, 5-yr.(% per annum)
92	FYGT10	Δlv	Interest rate: u.s. Treasury const maturities, 10-yr.(% per annum)
93	FYAAAC	Δlv	Bond yield: Moody's AAA corporate (% per annum)
94	FYBAAC	Δlv	Bond yield: Moody's BAA corporate (% per annum)
95	SCP90	lv	CP90 – FYFF (spread)
96	SFYGM3	lv	FYGM3 – FYFF (spread)
97	SFYGM6	lv	FYGM6 – FYFF (spread)
98	SFYGT1	lv	FYGT1 – FYFF (spread)
99	SFYGT5	lv	FYGT5 – FYFF (spread)
100	SFYGT10	lv	FYGT10 – FYFF (spread)
101	SFYAAAC	lv	FYAAAC – FYFF (spread)
102	SFYBAAC	lv	FYBAAC – FYFF (spread)
103	EXRUS	$\Delta \ln$	United States; effective exchange rate (merm) (index no.)
104	EXRSW	$\Delta \ln$	Foreign exchange rate: Switzerland (Swiss franc per U.S.$)
105	EXRJAN	$\Delta \ln$	Foreign exchange rate: Japan (yen per U.S.$)
106	EXRUK	$\Delta \ln$	Foreign exchange rate: United Kingdom (cents per pound)
107	EXRCAN	$\Delta \ln$	Foreign exchange rate: Canada (Canadian$ per U.S.$)

#	Mnemonic	Trans	Description
108	PWFSA	$\Delta^2 \ln$	Producer price index: finished goods (82=100,sa)
109	PWFCSA	$\Delta^2 \ln$	Producer price index: finished consumer goods (82=100,sa)
110	PWIMSA	$\Delta^2 \ln$	Producer price index: intermed mat. supplies & components (82=100,sa)
111	PWCMSA	$\Delta^2 \ln$	Producer price index: crude materials (82=100,sa)
112	PSCCOM	$\Delta^2 \ln$	Spot market price index: bls & crb: all commodities(1967=100)
113	PSM99Q	$\Delta^2 \ln$	Index of sensitive materials prices (1990=100)(bci-99a)
114	PMCP	lv	NAPM commodity prices index (percent)
115	PUNEW	$\Delta^2 \ln$	CPI-u: all items (82-84=100,sa)116
117	PU84	$\Delta^2 \ln$	CPI-u: transportation (82-84=100,sa)
118	PU85	$\Delta^2 \ln$	CPI-u: medical care (82-84=100,sa)
119	PUC	$\Delta^2 \ln$	CPI-u: commodities (82-84=100,sa)
120	PUCD	$\Delta^2 \ln$	CPI-u: durables (82-84=100,sa)
121	PUS	$\Delta^2 \ln$	CPI-u: services (82-84=100,sa)
122	PUXF	$\Delta^2 \ln$	CPI-u: all items less food (82-84=100,sa)
123	PUXHS	$\Delta^2 \ln$	CPI-u: all items less shelter (82-84=100,sa)
124	PUXM	$\Delta^2 \ln$	CPI-u: all items less medical care (82-84=100,sa)
125	GMDC	$\Delta^2 \ln$	PCE, impl price deflator (1987=100)
126	GMDCD	$\Delta^2 \ln$	PCE, impl price deflator: Durables (1987=100)
127	GMDCN	$\Delta^2 \ln$	PCE, impl price deflator: Nondurables (1996=100)
128	GMDCS	$\Delta^2 \ln$	PCE, impl price deflator: Services (1987=100)
129	CES275	$\Delta^2 \ln$	Average hourly earnings of production or nonsupervisory workers on private nonfarm payrolls: goods
130	CES277	$\Delta^2 \ln$	Average hourly earnings of production or nonsupervisory workers on private nonfarm payrolls: construction
131	CES278	$\Delta^2 \ln$	Average hourly earnings of production or nonsupervisory workers on private nonfarm payrolls: manufacturing
132	HHSNTN	Δlv	U. of Michigan index of consumer expectations

Table C1. Average Posterior Probabilities of Explanatory Variables in the 3-variable VAR

Explanatory variables	u_{t+12}	cpi_{t+12}	r_{t+12}
Personal income	0.141	0.001	0.949
IP index - Final products	0.251	0.003	0.564
IP index - Manufacturing	0.593	0.016	0.17
Capacity Utilization	1	0.124	0.032
Employment ratio	0.011	0.002	0.992
Civilian labor force: Total employed	0.428	0.003	0.652
Employees on nonfarm payrolls - Total private	0.811	0.018	0.317
Employees on nonfarm payrolls - Manufacturing	0.5	0.014	0.33
Employees on nonfarm payrolls - Service-providing	1	0.023	0.826
Employees on nfm prl - Trade, transportation and utilities	0.878	0.003	0.682
Employees on nonfarm payrolls - Wholesale trade	0.296	0.003	1
Employees on nonfarm payrolls - Financial activities	0.687	0.008	0.697
Average weekly hours of production	0.001	0.082	0.941
Housing starts: Total	0.879	0.001	0.04
Housing authorized: Total	1	0.001	1
Houses authorized by building permits: Northeast	1	0.105	0.003
Houses authorized by building permits: Midwest	1	0.025	0.018
Houses authorized by building permits: South	1	0.001	0.006
Houses authorized by building permits: West	1	0	1
Consumer installment credit to Personal income (ratio)	0.013	0.001	1
S&P'S common stock price index: Composite	0.962	0.132	0.004
S&P's composite common stock: Dividend yield	0.092	0.001	0.937
Commercial paper rate (spread from Fed Funds Rate)	0.028	0.7452	0.851
3-month interest rate (spread from FFR)	0.002	0.087	1
6-month interest rate (spread from FFR)	0.005	0.002	1
1-year interest rate (spread from FFR)	0.941	0.752	0.992
5-year interest rate (spread from FFR)	1	0.982	1
10-year interest rate (spread from FFR)	1	0.861	1
Bond yield: Moody's BAA corporate (spread from FFR)	0.001	0	0.978
NAPM commodity prices index	0.0012	0.867	0.857
CPI-U: Durables	0.172	0.002	0.543
CPI-U: All items less shelter	0.246	0.006	0.692

Table C2. Average Posterior Probabilities of AR-lags in the 3-variable VAR

Dependent Variable	Most important lags (probability>0.5)	Average posterior probability
u_{t+12}	r_{t-7}	0.56
	r_{t-7}	0.74
cpi_{t+12}	Own lags 1 to 7 (i.e. cpi_t to cpi_{t-6})	1
	cpi_{t-7}	0.83
r_{t+12}	r_{t-6}	1

Table C3. Forecast Comparison - relative RMSFE

	PI	IP	EMP	UR	TBILL	PPI	CPI	PCED
BVAR with factors (Bayesian Model Averaging)								
$h = 1$	0.94	1	0.9	0.96	1.08	0.88	0.95	1.09
$h = 4$	1.06	0.96	0.93	0.94	0.95	0.92	1.05	0.94
$h = 12$	0.97	0.92	0.99	1.02	0.98	0.92	0.95	0.96
BVAR with factors (Model Selection)								
$h = 1$	0.86	0.98	0.87	0.96	1.06	0.91	0.93	0.91
$h = 4$	0.9	0.97	0.85	0.92	0.94	0.94	0.98	0.93
$h = 12$	0.87	0.99	0.91	0.98	0.89	0.87	0.99	0.96
VAR with factors (BIC Selection)								
$h = 1$	0.92	0.99	0.94	0.99	1.22	0.99	1.01	0.97
$h = 4$	0.93	0.97	0.94	0.94	1.12	0.97	1.06	0.94
$h = 12$	0.97	1.04	0.98	1.05	0.99	0.9	1.1	0.95

Note: The variables of interest are: PI: Personal Income (A0M052), IP: Industrial Production(IPS10), EMP:Employment Rate (CES002), UR: Unemployment Rate (LHUR), TBILL: 3-month Treasury Bill Rate (FYGM3), PPI: Producer Price Index (PWFSA), CPI: Consumer Price Index (PUNEW), andPCED: PCE Deflator (GMDC)