

Forecasting with Dimension Switching VARs

Gary Koop
University of Strathclyde

October 2012

Abstract: This paper develops methods for Bayesian VAR forecasting when the researcher is uncertain about which variables enter the VAR and the dimension of the VAR may be changing over time. It considers the case where there are N variables which might potentially enter a VAR and the researcher is interested in forecasting N^* of them. Thus, the researcher is faced with 2^{N-N^*} potential VARs. If N is large, conventional Bayesian methods can be infeasible due to the computational burden of dealing with a huge model space. Allowing for the dimension of the VAR to change over time only increases this burden. In light of these considerations, this paper uses computationally practical approximations adapted from the dynamic model averaging literature so as to develop methods for dynamic dimension selection (DDS) in VARs. In an inflation forecasting application, we show the benefits of DDS. In particular, DDS switches between different parsimonious VARs and forecasts appreciably better than various small and large dimensional VARs.

Keywords: Bayesian VAR, model selection, variable selection, predictive likelihood

Acknowledgements: I would like to thank Dimitris Korobilis for helpful discussions. Gary Koop is a Fellow of the Rimini Centre for Economic Analysis. This research was supported by the ESRC under grant RES-062-23-2646.

1 Introduction

Vector autoregressions (VARs) are among the most popular tools in modern empirical macroeconomics. Large theoretical and empirical literatures exist on Bayesian VAR forecasting. There are many modelling and specification choices one faces when working with VARs and various Bayesian methods have been developed for dealing with them. Examples of recent surveys or empirical papers investigating such choices include Koop and Korobilis (2009), Carriero, Clark and Marcellino (2011), Del Negro and Schorfheide (2011) and Karlsson (2012). However, one important aspect of specification choice has been relatively neglected in the Bayesian VAR literature. This is the choice of the dimension of the VAR (i.e. which variables to include as dependent variables in the VAR).

Issues relating to the dimension of a VAR issue have increased in importance due to the growth of the large VAR literature. Papers such as Banbura, Giannone and Reichlin (2010), Carriero, Clark and Marcellino (2011), Carriero, Kapetanios and Marcellino (2009), Giannone, Lenza, Momferatou and Onorante (2010) and Koop (2011) work with VARs that have tens or even over a hundred dependent variables. In this literature, it is common for the researcher to be interested in forecasting a small number of variables (e.g. inflation or unemployment). There are many other variables which are potentially useful for forecasting and, if so, they should be included in the VAR. However, most of these other potential variables might be irrelevant and omitting them would lead to a more parsimonious VAR and improved forecasts. The problem is that the researcher does not know, a priori, which of these extra variables should be included. This problem is addressed in the large VAR literature by simply including all the potential dependent variables, but using an informative prior to shrink their effects so as to avoid over-fitting. It is worth emphasizing that this shrinkage is done on the VAR coefficients, but the dimension of the VAR always remains the same. Other approaches to VAR variable selection, such as Korobilis (2012), also focus on the choice of explanatory variables (lagged dependent variables), but maintain the full set of dependent variables at all points in time.

An alternative to shrinking coefficients is to develop statistical methods for selecting the appropriate dimension of the VAR. This is the challenge taken up in the small Bayesian literature on VAR dimension selection (see, e.g., Andersson and Karlsson, 2009, Jarocinski and Mackowiak, 2011 and Ding and Karlsson, 2012). One contribution of the present paper is to add to this literature, developing a new method for VAR dimension selection. However, the major contribution of this paper lies in doing dimension selection in a time-varying manner. That is, our method allows for the dimension of the VAR to change over time. Thus, the forecasting model may switch from, e.g., a small VAR to a larger VAR as the relevant set of forecasting variables switches over time. When forecasting a particular variable or set of variables, other predictors can enter and leave the VAR in a data-based manner so as to improve forecasting performance. For instance, our application involves forecasting inflation. Our approach allows a univariate AR model to be used to forecast inflation at some points in time, bivariate VARs (e.g. involving inflation and another predictor such as unemployment or the oil price) at other times and n -variate VARs at other times. We allow for every possible combinations of up to ten dimensional VARs (i.e. $n = 1, 2, 3, \dots, 10$). Thus, if N is the maximum VAR dimension and there are N^* variables we are interested in

forecasting, we are choosing between 2^{N-N^*} VARs at each point in time.

Allowing for such dimension switching has, to our knowledge, only been addressed in the Bayesian literature by Koop and Korobilis (2012) in a different context (with time-varying parameters in the VAR) and a vastly reduced model space involving 3 VARs instead of the 2^{N-N^*} VARs considered in the present paper. Here we are faced with a huge model space, unless N is very small or N^* is very close to N . This huge model space means that even standard Bayesian methods which do not allow for dimension switching will be computationally burdensome or infeasible. For instance, simply calculating marginal likelihoods for 2^{N-N^*} VARs will be computationally daunting even if one is working with a homoskedastic VAR with natural conjugate prior. Allowing for empirically important extensions such as heteroskedasticity will increase this burden. Further allowing for VAR dimension switching would add even greater complications. In light of these computational restrictions, we implement VAR dimension switching using an approximate method. This approximate method extends and adapts the dynamic model averaging (DMA) methodology of Raftery, Karny and Ettlter (2010). This approach was developed for time-varying parameter (TVP) regression models, but we adapt it for VARs. A key component in this approach is the predictive likelihood (i.e. the predictive density for the dependent variables evaluated at the observed outcome). With VARs of different dimensions, the predictive likelihoods are not comparable since the different VARs have different vectors of dependent variables. To surmount this problem, we adopt a strategy used in Andersson and Karlsson (2009) and Ding and Karlsson (2012) and use the predictive likelihood for the dependent variables which are common to all models (i.e. the N^* variables we are interested in forecasting). It is also worth noting that DMA, as its name suggests, is a method for model averaging. However, it can also be used for model selection and, in this paper, we use it in this sense (although it is straightforward to adapt our approach to do model averaging). Thus, we use the terminology dynamic dimension selection (DDS) to describe our methodology which allows for VARs of different dimension to be selected in a time-varying manner.

In an inflation forecasting exercise, we find our DDS methodology to forecast better than several standard, fixed-dimensional, VAR approaches. We show how substantial dimension switching does occur and precisely which variable(s) enter/leave the selected inflation forecasting model as time evolves.

2 The Econometrics of Bayesian VARs

Let y_t be an N -vector containing all the potential dependent variables in the VAR. Our model space is defined through the following set of VARs:

$$y_t^{(m)} = Z_t^{(m)} \beta^{(m)} + \varepsilon_t^{(m)},$$

for $t = 1, \dots, T$ where $\varepsilon_t^{(m)}$ is i.i.d. $N\left(0, \Sigma_t^{(m)}\right)$, $y_t^{(m)}$ is an $n \times 1$ vector containing n of the N variables in y_t ,

$$Z_t^{(m)} = \begin{pmatrix} z_t^{(m)} & 0 & \dots & 0 \\ 0 & z_t^{(m)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & z_t^{(n)} \end{pmatrix},$$

and $z_t^{(m)}$ is a row vector containing an intercept and p lags of each of the n variables in $y_t^{(m)}$.

Our model space is defined by these $m = 1, \dots, M$ VARs. We divide the dependent variables into a set of N^* variables that we are interested in forecasting, y_t^f , and the remainder, y_t^r . $y_t^{(m)}$ always includes y_t^f and the different models are defined by different subsets of y_t^r . Since there are 2^{N-N^*} possible subsets of y_t^r , we have $M = 2^{N-N^*}$ VARs.

Analytical results exist for posterior and predictive analysis with homoskedastic Bayesian VARs when a natural conjugate prior is used. However, in many applications it is important to allow the errors to be heteroskedastic. When heteroskedasticity is present, analytic posterior and predictive results are lost and an exact Bayesian analysis requires the use of computationally demanding MCMC methods. With large model spaces, doing MCMC in every model can be computationally infeasible. However, if $\Sigma_t^{(m)}$ is a known matrix, analytical results are again available. For this reason, we replace $\Sigma_t^{(m)}$ with an estimate. In particular, we use an Exponentially Weighted Moving Average (EWMA) estimate to model volatility (see RiskMetrics, 1996 and West and Harrison, 1997):

$$\widehat{\Sigma}_t^{(m)} = \kappa \widehat{\Sigma}_{t-1}^{(m)} + (1 - \kappa) \widehat{\varepsilon}_t^{(m)} \widehat{\varepsilon}_t^{(m)'} \tag{1}$$

where $\widehat{\varepsilon}_t = y_t - Z_t^{(m)} E\left(\beta^{(m)} | Data_t\right)$ where $Data_t$ is data available through time t and $E\left(\beta^{(m)} | Data_t\right)$ is the posterior mean of $\beta^{(m)}$.¹ EWMA estimators require the selection of the decay factor, κ . In our empirical work, we work with various values of the decay factor in the interval $[0.90, 0.99]$.

We also consider a homoskedastic variant of our VARs. In this case, we use a standard recursive estimate. When forecasting using data through time t , we estimate $\Sigma^{(m)}$ as

$$\widehat{\Sigma}^{(m)} = \frac{\sum_{i=1}^t \widehat{\varepsilon}_i^{(m)} \widehat{\varepsilon}_i^{(m)'}}{t}.$$

With Σ_t being treated as fixed, a normal prior for $\beta^{(m)}$ will lead to an analytical posterior and one-step ahead predictive density. We use a version of the popular Minnesota prior (see Doan, Litterman and Sims, 1984 and Litterman, 1986) similar to that used in Banbura, Giannoni and Reichlin (2010) and Carriero, Clark and Marcellino (2011) and many other recent papers. This prior depends on one shrinkage parameter,

¹In our empirical work, the initial condition, $\widehat{\Sigma}_0$, is set to be the sample covariance matrix using an initial $\tau = 48$ months of data. Our forecast evaluation begins in month $\tau + 1$.

λ ,² and is given by

$$\beta^{(m)} \sim N\left(\underline{\beta}^{(m)}, \underline{V}^{(m)}\right). \quad (2)$$

We shrink towards random walks by setting all elements of $\underline{\beta}^{(m)}$ to zero except for the elements corresponding to the first own lag of the dependent variable in each equation. These elements are set to one. The degree of shrinkage is controlled by $\underline{V}^{(m)}$ which is assumed to be a diagonal matrix. The elements corresponding to the intercepts in each equation are set to large numbers (i.e. a noninformative prior is used for the intercepts). All other diagonal elements are set to $\frac{\lambda}{r^2}$ for $r = 1, \dots, p$. That is, the prior variance on the r^{th} lag of a dependent variable is $\frac{\lambda}{r^2}$ thus ensuring greater shrinkage on more distant lags. The shrinkage parameter, λ , is estimated from the data as described below.³ The precise formulae for posterior and one-step ahead predictive density are available in many sources (e.g. Koop and Korobilis, 2009) and will not be reproduced here.

3 Model Switching Using Forgetting Factors

The aim of this paper is to do model selection in a time-varying manner. To do this, we adapt methods developed in Raftery et al (2010) to our VAR context. Raftery et al (2010) refer to their methods as DMA, although the methods can also be used for dynamic model selection (DMS) where a different model may be selected as the forecasting model at each point in time. Raftery et al (2010) provide a complete derivation and explanations of the attractions of DMA and how it can over-come the computational burdens associated with a high dimensional model space. The reader is referred to their paper for such details and only a practical summary is provided here.

The goal of DMS is to calculate $\pi_{t|t-1,m}$ for the $m = 1, \dots, M$ VARs. $\pi_{t|t-1,m}$ is the probability that model m is the forecasting model at time t given data available at time $t - 1$. DMS involves selecting the model with highest value for $\pi_{t|t-1,m}$ at each point in time and using it to forecast y_t^f . Since $\pi_{t|t-1,m}$ changes over time, the forecasting model can switch over time. Calculating $\pi_{t|t-1,m}$ using a fully defined Bayesian model (e.g. involving a Markov transition matrix to model the switches between forecasting models) will typically lead to over-parameterization worries and be computationally burdensome unless the number of models is small. But Raftery et al (2010) develop a fast recursive algorithm for calculating $\pi_{t|t-1,m}$ under an approximation involving a so-called forgetting factor.

A key ingredient in this algorithm is the predictive likelihood, which is a measure of forecast performance. We denote the predictive density for the variables being forecast (y^f) by $p_m\left(y_t^f | Data_{t-1}\right)$. The formula for the predictive density for the VAR (with normal prior and Σ_t replaced by an estimate) is available in standard sources such as Koop and Korobilis (2009). The predictive likelihood is this predictive density evaluated at the actual realization which we denote by \tilde{y}_t^f . We stress that, following Andersson and

²The extension to multiple shrinkage parameters is theoretically straightforward, but would increase the computational burden.

³The slight difference in prior formulation from, e.g., Carriero, Clark and Marcellino (2011) is due to the fact that we standardize all of our variables. This standardization is done by subtracting off the mean and dividing by the standard deviation calculated using the first $\tau = 48$ observations.

Karlsson (2009) and Ding and Karlsson (2012), we are using the predictive likelihood, $p_m \left(y_t^f = \tilde{y}_t^f | Data_{t-1} \right)$, for variables which are common to every model.

The algorithm begins by specifying initial model probabilities $\pi_{0|0,m}$ for $m = 1, \dots, M$. In our empirical work we adopt the noninformative choice $\pi_{0|0,m} = \frac{1}{M}$. It then uses a model prediction equation using a forgetting factor α :

$$\pi_{t|t-1,m} = \frac{\pi_{t-1|t-1,m}^\alpha}{\sum_{j=1}^M \pi_{t-1|t-1,j}^\alpha}, \quad (3)$$

and a model updating equation of:

$$\pi_{t|t,m} = \frac{\pi_{t|t-1,m} p_m \left(y_t^f = \tilde{y}_t^f | Data_{t-1} \right)}{\sum_{l=1}^M \pi_{t|t-1,l} p_l \left(y_t^f = \tilde{y}_t^f | Data_{t-1} \right)}. \quad (4)$$

Thus, the algorithm is a simple filtering algorithm which (provided $p_m \left(y_t^f = \tilde{y}_t^f | Data_{t-1} \right)$ can be evaluated analytically, as it the case in our paper) is computationally fast and efficient.

The use of forgetting factors in order to simplify the recursions in state space models has a long history (see, e.g., Fagin, 1964, and Jazwinsky, 1970) and their properties are well understood. The contribution of Raftery et al (2010) arose from using a forgetting factor approach for model averaging/switching. They provide additional details, including several ways of justifying their approach. One justification of their approach can be seen by noting that it implies:

$$\pi_{t|t-1,m} \propto \prod_{i=1}^{t-1} \left[p_m \left(y_i^f = \tilde{y}_i^f | Data_{i-1} \right) \right]^\alpha.$$

In words, the algorithm allocates more probability to models which have forecast well (as measured by the predictive likelihood) in the past. The forgetting factor, α , controls the rate at which past forecasting performance is weighted. If $\alpha = 1$, then forecasting performance from all past periods from $i = 1, \dots, t-1$ are equally weighted when deciding on a forecasting model for time t . If $\alpha = 0.95$, then forecast performance 12 months ago receives (roughly) half as much weight as last period's forecasts. In our empirical work, we consider $\alpha = 1$ (which allows for very slow model switching), $\alpha = 0.99$ (slow model switching), $\alpha = 0.95$ (moderate model switching) and $\alpha = 0.90$ (rapid model switching). The choice $\alpha = 1$ will be equivalent to Bayesian model averaging (BMA) done in a recursive manner with an expanding window of data.

The theory outlined in this section holds for any set of models. In Raftery et al (2010) the different models were TVP regression models. In the present paper, the different models are VARs. But the same algorithm can be used.

In addition, for the Bayesian a model also involves a prior. Thus, we can augment our model space by considering different priors. Remember that our Minnesota prior depends on a shrinkage parameter, λ . We consider five different values for λ and interpret these as five different models. In particular, we let $\lambda = 10^{-10}, 0.001, 0.005, 0.01, 10$. These values vary from an extremely tight prior implying very strong shrinkage towards a random walk to very noninformative priors imply minimal shrinkage. At each point

in time, the optimal value of λ is selected using DMS. Thus, our entire model space contains five versions of each of the n -variate VARs and we are estimating the Minnesota prior shrinkage parameter in a time-varying manner.

In summary, we are using an existing algorithm which allows for model switching. However, we are using it for a different set of models than has been considered in the past. Our set of models differ in VAR dimensionality and the degree of shrinkage in the Minnesota prior. Since the main focus is on VAR dimension selection, we use the terminology DDS for this approach.

4 Empirical Results

In this section we investigate the ability of our DDS methodology to forecast US inflation using up to 10 dependent variables in a VAR. It is divided into three sub-sections which: i) describe the data, ii) investigate which VAR dimensions and which variables DDS selects and iii) compare the forecast performance of DDS relative to other VAR-based methods.

4.1 Data

We use data on 10 commonly-used monthly US macroeconomic variables from 1973M1 through 2012M3. All of the data is obtained from the Federal Reserve Bank of St. Louis' FRED data base. We follow a common practice (see, e.g., Carriero, Clark and Marcellino, 2011) and transform our variables so as to be rates. Table 1 lists our variables and the transformations done. Variables which were available weekly or daily are averaged to produce monthly figures. Where relevant, variables are seasonally adjusted. The results below maintain a lag length of 4, but results are relatively insensitive to lag length. Inflation is the variable common to all VARs (labelled y_t^f in Section 2), with the remaining variables being potential predictors for inflation (labelled y_t^r in Section 2).

Acronym	Original Variable	Transformation
INF	PCE price index	log difference
UR	Unemployment rate	none
CLAIMS	Initial unemployment claims	none
IP	Industrial production	log difference
ORDERS	ISM manuf. new orders index	none
POIL	WTI spot oil price	log difference
SP500	S&P500 index	log difference
ILONG	10-year t-bill rate	none
ISHORT	Federal funds rate	none
EXR	Trade-weighted US \$ exchange rate	log difference

4.2 Which Variables Does DDS Select?

The two main specification choices in DDS are α (which affects the degree of model switching) and κ (which affects the amount of change in Σ_t). In the following section,

a forecasting exercise will be presented which investigates a wide range of values for α and κ . However, before doing this, we present evidence on what DDS is doing for the particular choices $\alpha = 0.99$ and $\kappa = 0.90$. The value $\alpha = 0.99$ is the one suggested by Raftery et al (2010) and $\kappa = 0.90$ is a value which allows for a high degree of variation in volatility. As we shall see in the next sub-section, allowing for substantive heteroskedasticity does seem to be important in achieving good forecast performance in this data set.

Figure 1 shows the size of the VAR which is chosen by DDS at each point in time. The two main points to note are that DDS is ensuring parsimony and that the VAR dimension switches substantially over time.

On the issue of parsimony, it can be seen that DDS is never choosing a VAR with more than six variables and for most of the time, the dimension is three or less. In fact, for the latter half of the 1970s it is selecting a univariate AR model. Subsequent to the 1970s, the VAR dimension tended to increase in size until an abrupt switch in dimension around 1990. For a long period from 1990 through 2008, DDS chose a three dimensional VAR (occasionally and briefly switching to a two or four dimensional model). It is only in the recent recessionary period that the VAR dimension has been increasing.

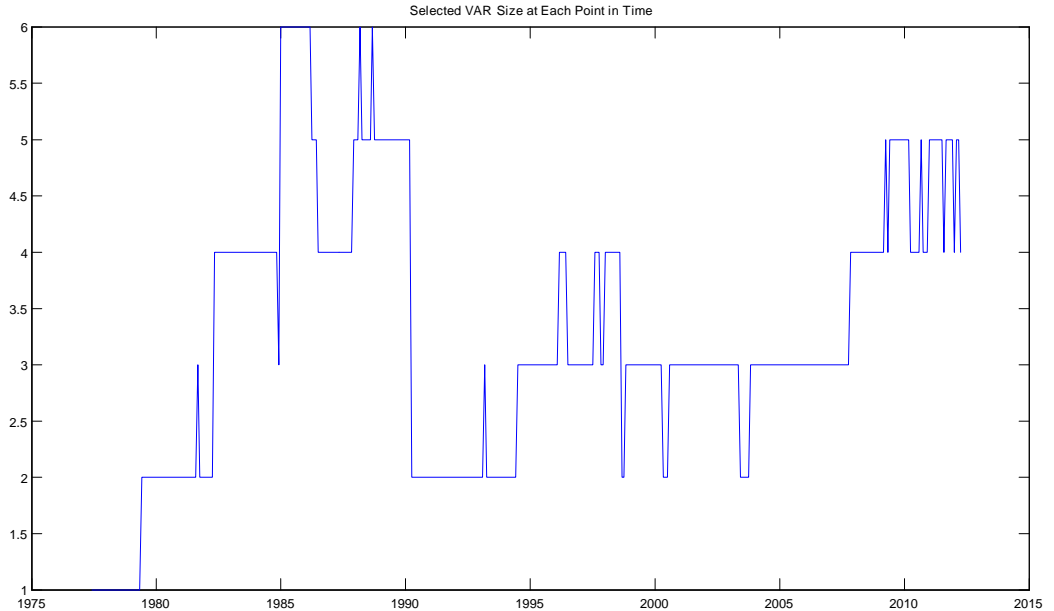


Figure 1

Figures 2 and 3 show which variables DDS has been selecting at each point in time. The numbers in the figures are “inclusion indicators” which equal one in time periods when DDS is selecting the variable for inclusion in the VAR (and equal zero at other times). The dominant variable is oil price inflation which, apart from the late 1970s and a few other brief periods, is selected by DDS. During the period 1990-2008, when DDS

is choosing parsimonious VARs, it is typically using a trivariate VAR involving INF, POIL and SP500. After 2000 (and, in particular, after 2008 when the VAR dimension tends to increase), the exchange rate and unemployment rate also frequently enter as important variables. It is interesting to note that, in the 1980s (when DDS selects VARs with higher dimensions) the long and short term interest rates are included, but subsequently these variables play no role whatsoever. The remaining variables (ORDERS, CLAIMS, IP) only occasionally enter the VAR. Although it is interesting to note that each of these variables is selected at some point in time, indicating they are episodically useful for forecasting inflation.

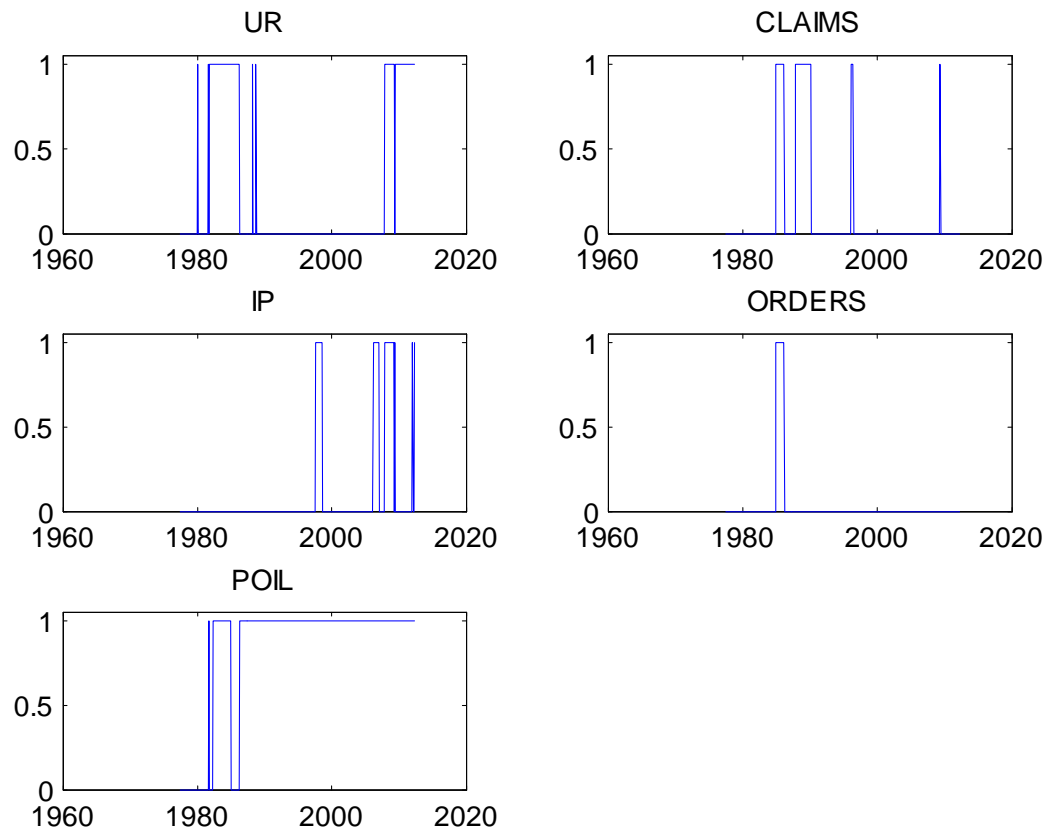


Figure 2: Inclusion Indicators for Individual Variables

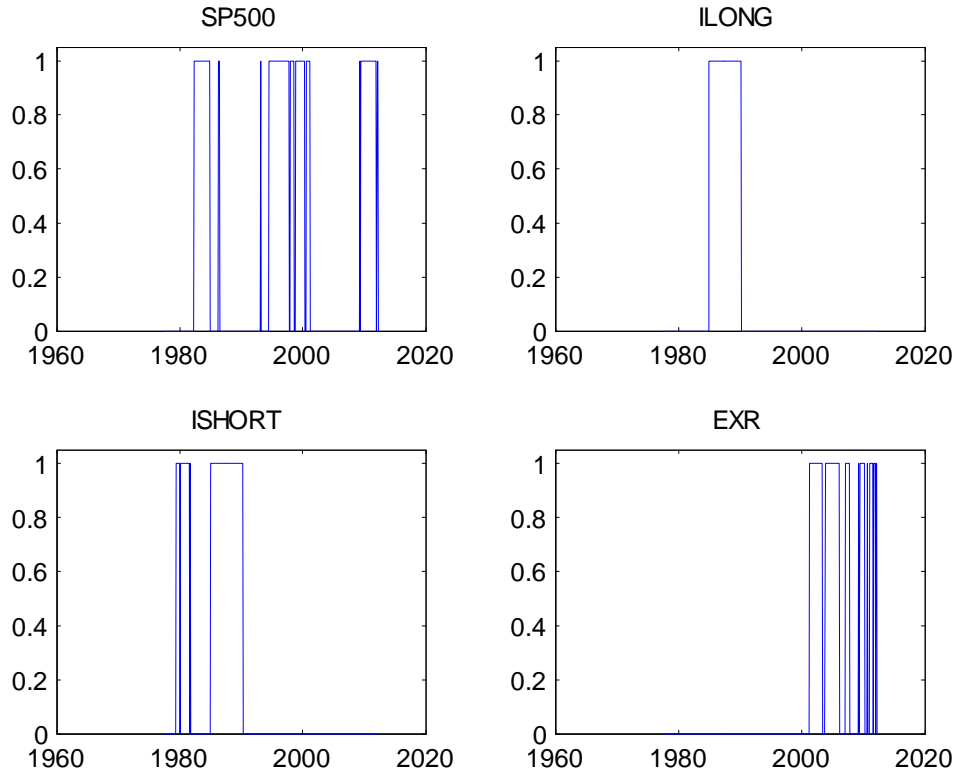


Figure 3: Inclusion Indicators for Individual Variables

4.3 Forecasting Comparison

The preceding sub-section indicates that the VAR dimension is switching over time and that, when forecasting inflation, different variables are useful at different points in time. In the present sub-section, we investigate whether this switching is leading to substantial improvements in forecast performance. Given that DDS uses the predictive likelihood which evaluates the one-period ahead forecast performance, we focus on one-period forecasts in this sub-section.

We compare DDS forecast performance to that of several forecasting strategies a researcher might use with VARs of fixed dimensions. In particular, we reason that a researcher might choose one of three strategies: i) use the large VAR (include all 10 variables), ii) motivated by the Phillips curve, use a bivariate VAR for inflation and unemployment, or iii) use a univariate AR for inflation. In order to make sure our forecast comparison is as fair as possible, we use the same Minnesota priors and same strategy for optimizing the shrinkage parameter in the Minnesota prior in the fixed dimension VARs as we did with DDS. We also allow for heteroskedasticity using the same EWMA estimator as we did with DDS. Homoskedastic variants of the VARs are also treated in the same way for both cases.

We use three standard metrics of forecast performance. Mean squared forecast error (MSFE) and mean absolute forecast error (MAFE) investigate the properties of

the point forecasts of inflation. Sums of log predictive likelihoods (LPL) can be used to investigate the forecast performance of the entire predictive density.⁴ All of these are evaluated over the period 1977M1 through 2012M3

We begin by investigating the sensitivity of DDS to choice of forgetting and decay factors. A strong finding of Table 2 is that allowing for changes in volatility is important in achieving good forecast performance. Imposing homoskedasticity, or even $\kappa = 0.99$, leads to inferior forecast performance relative to values of the decay factor which allow for greater volatility change. This result comes through with moderate strength in the MSFEs and MAFEs. However, it comes through with greater strength in the LPLs which depend not only on the point forecast, but also the predictive variance. This indicates that appropriate modelling of heteroskedasticity is of particular importance in correctly estimating the dispersion of the predictive distribution.

If we ignore the poorly performing homoskedastic and $\kappa = 0.99$ rows of Table 2, it can be seen that MSFE and MAFE results are fairly robust to the choice of α . Values of α allowing for a moderate degree of model switching (i.e. $\alpha = 0.95$ or 0.99) tend to perform best, with $\alpha = 0.90$ performing worst. The combination $\alpha = 0.99$ and $\kappa = 0.90$ leads to the best forecast performance as evaluated by LPL. If we use MSFE or MAFE, the combination $\alpha = 0.95$ and $\kappa = 0.95$ forecasts slightly better. But, in general, we are finding a fair degree of robustness provided we avoid the homoskedastic case and do not allow for too much model switching. Note that, for $\alpha = 0.90$, forecasts one year ago only receive about 28% as much weight as last month's forecast performance in the model switching choice. Thus, this value is choosing the forecasting model based only on the very recent past, which is not optimal in the present application.

α	κ	LPL	MSFE	MAFE
1.00	Homo	-490.53	0.611	0.590
1.00	0.99	-468.14	0.564	0.571
1.00	0.95	-428.00	0.565	0.562
1.00	0.90	-408.92	0.566	0.565
0.99	Homo	-489.04	0.617	0.590
0.99	0.99	-467.45	0.570	0.567
0.99	0.95	-417.58	0.550	0.553
0.99	0.90	-397.90	0.550	0.556
0.95	Homo	-497.09	0.632	0.594
0.95	0.99	-473.81	0.608	0.579
0.95	0.95	-412.18	0.537	0.542
0.95	0.90	-402.66	0.588	0.568
0.90	Homo	-499.20	0.634	0.600
0.90	0.99	-483.87	0.642	0.586
0.90	0.95	-424.36	0.585	0.567
0.90	0.90	-409.43	0.597	0.579

⁴To aid in interpretation, note that sums of log predictive likelihoods over the entire sample will equal the log marginal likelihood. The log Bayes factor is equivalent to the difference in the log marginal likelihoods between two models. There are standard rules of thumb for interpreting Bayes factors (see, e.g., Kass and Raftery, 1995). Also, information criteria such as the Schwarz criterion can be interpreted as approximations to log marginal likelihoods.

Table 3 presents forecast metrics for the fixed dimension VARs for different values for κ . As in Table 2, a strong finding is that allowing for substantial change in volatility is important for achieving good forecast performance as measured by LPL (results for MSFE and MAFE are weaker). However, the most important finding is that the forecast performance of fixed dimension VARs tends to be substantially worse than DDS. For instance, the MSFEs in Table 3 are rarely much below 0.60. In Table 2, if we ignore the poorly forecasting homoskedastic results, we tend to find MSFE values of around 0.55-0.57 (and the lowest MSFE using DDS is 0.537). This represents a 5-10% improvement in forecast performance which, for inflation, is substantial. Similarly, in Table 3 the highest LPL is -416.23 which is much lower than the highest LPL of Table 2 which is -397.90. In general, if we compare like with like (e.g. compare $\kappa = 0.95$ rows in the two tables), you see a deterioration in forecast performance when moving from DDS to the fixed dimension VAR.

The final row of Table 3 presents MSFE and MAFE for a no change forecast. These, too, are substantively higher than the ones produced by DDS.

It is also interesting to note that, with fixed dimension VARs, LPLs indicate that the univariate AR model for inflation is forecasting best. In contrast, MSFEs and MAFEs show no strong preference for any particular VAR dimension. The fact that DDS is forecasting better (by any of our forecast metrics) and is selecting parsimonious (but rarely univariate) VARs indicate the importance of dimension switching and allowing for different predictors for inflation to enter/leave the VAR over time.

VAR Dimension	κ	LPL	MSFE	MAFE
10	Homo	-475.66	0.590	0.579
10	0.99	-473.28	0.586	0.579
10	0.95	-448.46	0.599	0.584
10	0.90	-433.89	0.612	0.592
2	Homo	-485.07	0.605	0.584
2	0.99	-476.05	0.599	0.580
2	0.95	-440.97	0.598	0.578
2	0.90	-419.80	0.596	0.578
1	Homo	-482.54	0.599	0.576
1	0.99	-471.52	0.587	0.576
1	0.95	-437.15	0.594	0.578
1	0.90	-416.23	0.592	0.577
No change	—	—	0.724	0.636

These empirical results show the computational feasibility and usefulness of DDS in a VAR of up to 10 dimensions on a standard personal computer. Extensions to VARs somewhat larger than this (e.g. 15 or 20 dimensions) could be handled if a faster computer were used or the researcher was willing to tolerate a greater computation time. However, in the near future, it is unlikely that DDS as implemented in this paper could be done using large VARs involving 100 or more variables (such as the one in Banbura et al, 2010). Even a very fast computer would take many years to evaluate 2^{50} or 2^{100} VARs. In such a case, we would recommend restricting the model space in some way. For instance, the researcher might only work with all VARs up to a small fixed

dimension. Or the researcher may wish to divide variables into blocks (e.g. a financial block, a monetary block, etc.) and consider only VARs which select one variable from each block.

In our empirical work, the goal was to investigate the performance of DDS for a range of forgetting and decay factors. It is worth noting that it is also possible (at some computational cost) to estimate these factors. The interested reader is referred to McCormick, Raftery, Madigan and Burd (2012) and Koop and Korobilis (2012).

5 Conclusions

When working with Bayesian VARs, parsimony is often a concern. When forecasting, the researcher is often concerned with whether the set of variables useful for forecasting is changing over time. In this paper, we have developed DDS methods which address both these concerns. DDS can be used to choose VAR dimension and, in addition, allows the dimension of the VAR to switch over time. The results of an inflation forecasting exercise suggest that DDS is a potentially useful addition to the Bayesian VAR toolbox.

References

- Andersson, M. and Karlsson, S. (2009). “Bayesian forecast combination for VAR models,” pages 501-524 in S. Chib, W. Griffiths, G. Koop and D. Terrell (eds.), *Advances in Econometrics, Volume 23*, Emerald Group: Bingley, UK.
- Banbura, M., Giannone, D. and Reichlin, L. (2010). “Large Bayesian vector autoregressions,” *Journal of Applied Econometrics*, 25, 71-92.
- Carriero, A., Clark, T. and Marcellino, M. (2011). “Bayesian VARs: Specification choices and forecast accuracy,” Working Paper 1112, Federal Reserve Bank of Cleveland.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2009). “Forecasting exchange rates with a large Bayesian VAR,” *International Journal of Forecasting*, 25, 400-417.
- Del Negro, M. and Schorfheide, F. (2011). “Bayesian macroeconometrics,” pages 293-398 in J. Geweke, G. Koop, and H. van Dijk (eds.), *The Oxford Handbook of Bayesian Econometrics*, Oxford University Press: Oxford.
- Ding, S. and Karlsson, S. (2012). “Model averaging and variable selection in VAR models,” manuscript.
- Doan, T., Litterman, R. and Sims, C. (1984). “Forecasting and conditional projection using realistic prior distributions,” *Econometric Reviews*, 3, 1-144.
- Fagin, S. (1964). “Recursive linear regression theory, optimal filter theory, and error analyses of optimal Systems,” *IEEE International Convention Record Part i*, 216-240.
- Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2010). “Short-term inflation projections: a Bayesian vector autoregressive approach,” ECARES working paper 2010-011, Universite Libre de Bruxelles.
- Jarocinski, M. and Mackowiak, B. (2011). “Choice of variables in vector autoregressions,” manuscript.
- Jazwinsky, A. (1970). *Stochastic Processes and Filtering Theory*, New York: Academic Press.
- Karlsson, S. (2012). “Forecasting with Bayesian Vector Autoregressions,” prepared for the *Handbook of Economic Forecasting, Volume 2*.
- Kass, R. and Raftery, A. (1995). “Bayes factors,” *Journal of the American Statistical Association*, 90, 773-795.
- 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. (2012). “Large time-varying parameter VARs,” manuscript available through <http://personal.strath.ac.uk/gary.koop/research.htm>.
- Korobilis, D. (2012). “VAR forecasting using Bayesian variable selection”, *Journal of Applied Econometrics*, forthcoming.
- Litterman, R. (1986). “Forecasting with Bayesian vector autoregressions – Five years of experience,” *Journal of Business and Economic Statistics*, 4, 25-38.
- McCormick, T., Raftery, A., Madigan, D. and Burd, R. (2012). “Dynamic logistic regression and dynamic model averaging for binary classification,” *Biometrics*, 68, 23-30.
- Raftery, A., Karny, M. and Ettlér, P. (2010). “Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill,” *Technometrics*, 52, 52-66.
- RiskMetrics (1996). *Technical Document* (Fourth Edition). Available at <http://www.riskmetrics.com/system/files/private/td4e.pdf>.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, second edition, Springer: New York.