International Journal of Forecasting 30 (2014) 280–290

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting with dimension switching VARs

Gary Koop

University of Strathclyde, United Kingdom

ARTICLE INFO

Keywords:
Bayesian VAR
Model selection
Variable selection
Predictive likelihood

ABSTRACT

This paper develops methods for 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 in order to develop methods for dynamic dimension selection (DDS) in VARs. We then show the benefits of DDS in a macroeconomic forecasting application. In particular, DDS switches between different parsimonious VARs and forecasts appreciably better than various small and large dimensional VARs.

1. Introduction

Vector autoregressions (VARs) are among the most popular tools in modern empirical macroeconomics, and a large body of theoretical and empirical literature on Bayesian VAR forecasting exists. When working with VARs, one faces many modelling and specification choices, and various Bayesian methods have been developed for dealing with them. Examples of recent surveys or empirical papers investigating such choices include those of Carriero, Clark, and Marcellino (2011), Del Negro and Schorfheide (2011), Karlsson (2012) and Koop and Korobilis (2009). However, one important aspect of specification choice has been relatively neglected in the VAR literature: 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 have increased in importance recently, due to the growth of the large VAR literature. Papers such as those by Banbura, Giannone, and Reichlin (2010), Carriero et al. (2011), Carriero, Kapetanios, and Marcellino (2009), Giannone, Lenza, Momferatou, and Onorante (2010) and Koop (2013) work with VARs that have tens of dependent variables, or even over a hundred. In this literature, it is common for the researcher to be interested in forecasting a small number of variables (e.g., inflation and unemployment). There are many other variables which are potentially useful for forecasting, and any that prove useful should be included in the VAR. However, most of these potential variables turn out to be irrelevant, in which case omitting them will lead to a more parsimonious VAR and improved forecasts. The problem with this 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 of 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 that of Korobilis (2013), 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 the coefficients is to develop statistical methods for selecting the appropriate dimension of the VAR. This is the challenge which is taken

E-mail address: gary.koop@strath.ac.uk.

0169-2070/$ – see front matter
http://dx.doi.org/10.1016/j.ijforecast.2013.09.005

up in the small Bayesian literature on VAR dimension selection (see, e.g., Andersson & Karlsson, 2009; Ding & Karlsson, 2012; Jarocinski & Mackowiak, 2011). One contribution of the present paper is to add to this literature, developing a new method for VAR dimension selection. However, its major contribution lies in the fact that it performs dimension selection in a time-varying manner. That is, our method allows 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 the forecasting performance. Such an approach has been found to improve forecast performances in regressions (see e.g. Rossi & Sekhposyan, 2010). For instance, our application involves forecasting inflation, unemployment and industrial production. Our approach allows a trivariate VAR model using these variables to forecast inflation at some points in time, four-variable VARs (e.g., involving inflation, unemployment, industrial production and another predictor such as the oil price) at other times, and n-variate VARs at other times. We allow for every possible combination of up to 12 dimensional VARs (i.e., $n = 3, \ldots, 12$). 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.

To the best of our knowledge, allowing for such dimension switching has only been addressed in the Bayesian literature by Koop and Korobilis (2013), though in a different context (with time-varying parameters in the VAR) and a vastly reduced model space involving three 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 a 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 Ettler (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 by 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 that is the sense in which we use it in this paper (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 that substantial dimension switching does occur, and indicate 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 of 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 $t = 1, \ldots, T$, 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,

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

and $z_t^{(m)}$ is a row vector containing an intercept and lags of each of the n variables in $y_t^{(m)}$. When forecasting h periods ahead, $Z_t^{(m)}$ will contain information dated $t-h$ or earlier.

Our model space is defined by these $m = 1, \ldots, M$ 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, many macroeconomic applications have found that it is important to allow the errors to be heteroskedastic (e.g., Primiceri, 2005; and Sims & Zha, 2006). When heteroskedasticity is present, the 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. This has the drawback that it ignores parameter uncertainty relating to $\Sigma_t^{(m)}$, but the advantage of a great reduction in computational time, meaning that much larger model spaces can be handled. In particular, we use an Exponentially Weighted Moving Average (EWMA) estimate to model the volatility (see RiskMetrics, 1996; and