



**Koop, Gary (2017) Bayesian methods for empirical macroeconomics with big data. Review of Economic Analysis. ISSN 1973-3909 (In Press) ,**

This version is available at <http://strathprints.strath.ac.uk/60339/>

**Strathprints** is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: [strathprints@strath.ac.uk](mailto:strathprints@strath.ac.uk)

# Bayesian Methods for Empirical Macroeconomics with Big Data

Gary Koop

University of Strathclyde and Rimini Centre for Economic Analysis

**Abstract:** Bayesian econometric methods are increasingly popular in empirical macroeconomics. They have been particularly popular among macroeconomists working with Big Data (where the number of variables under study is large relative to the number of observations). This paper, which is based on a keynote address at the Rimini Centre for Economic Analysis' 2016 Money-Macro-Finance Workshop, explains why this is so. It discusses the problems that arise with conventional econometric methods and how Bayesian methods can successfully overcome them either through use of prior shrinkage or through model averaging. The discussion is kept at a relatively non-technical level, providing the main ideas underlying and motivation for the models and methods used. It begins with single-equation models (such as regression) with many explanatory variables, then moves on to multiple equation models (such as Vector Autoregressive, VAR, models) before tackling the challenge caused by parameter change (e.g. changes in VAR coefficients or volatility). It concludes with an example of how the Bayesian can address all these challenges in a large multi-country VAR involving 133 variables: 7 variables for each of 19 countries.

Keywords: multivariate time series, vector autoregression, state space model

JEL Classifications: C11, C32, C53

Acknowledgements: I would like to thank participants at the 7th RCEA Macro-Money-Finance Workshop on Advances in Macroeconomics and Finance for helpful comments.

# 1 Introduction

Big Data<sup>1</sup> has the potential to revolutionize empirical macroeconomics. The information contained in large data sets could improve our forecasts and our understanding of the macroeconomy. Big Data is available in many macroeconomic contexts. In most countries, government statistical agencies collect data on a wide range of macroeconomic variables (e.g. measures of output, capacity, employment and unemployment, prices, wages, housing, inventories and orders, stock prices, interest rates, exchange rates and monetary aggregates). In the US, the Federal Reserve Bank of St. Louis maintains the FRED-MD monthly data base for well over 100 macroeconomic variables from 1960 to the present (see McCracken and Ng, 2015). Many other countries have similar data sets. And, in an increasingly globalized world where economic developments in one country can affect others, the researcher may wish to work with data for several countries. Big Data can also arise through the wish to bring more and more financial variables, often at different frequencies, into a macroeconomic model.

In macroeconomics, the challenges which arise in the presence of Big Data are magnified by the fact that our models typically must include parameter change. For instance, the reduction of the volatilities of many macroeconomic variables for many countries, often known as the Great Moderation of the business cycle, which began in the early 1980s before being reversed by the Great Recession, means that econometric models should have time-varying error variances. There are often reasons for thinking coefficients in a regression or a Vector Autoregression (VAR) should also be time-varying. For instance, differences in the monetary policy transmission mechanism over the business cycle (e.g. printing money at the height of an expansion can have a very different impact on inflation than at the depth of a recession) imply coefficients in a VAR should be time-varying. Stories like this abound in macroeconomics to a greater extent than in other fields of economics and justify careful building of parameter change into models.

The interaction of Big Data with parameter change in macroeconomics raises challenges. In this paper I will elaborate on the nature of these challenges, why new methods are called for and describe some of them. I will adopt a largely non-technical approach, describing only the main ideas and providing references which point the reader in the direction of additional details. In this literature, Bayesian methods have proved particularly popular for reasons that will be made clear. Hence, I will focus on Bayesian methods in this paper. But it is worthwhile emphasizing that some of the methods I will discuss, such as the Least absolute shrinkage and selection operator (LASSO), can be interpreted either in a Bayesian or frequentist<sup>2</sup> fashion. And there do exist non-Bayesian methods for addressing the issues concerned (see, e.g., Hendry and Doornik, 2014).

Macroeconomists are interested in building econometric models for macroeconomic variables for many reasons including forecasting, estimating fully structural models (e.g. Dynamic Stochastic General Equilibrium or DSGE models) or calculating features of

---

<sup>1</sup>Big Data can come in two forms which Varian (2014) calls Tall and Fat. Tall Data arises when the number of observations becomes huge. Fat Data arises when the number of variables becomes huge. This paper is about Fat Data.

<sup>2</sup>Frequentist econometrics is the formal name given for conventional, non-Bayesian econometric methods based on the frequency theory of probability.

interest (e.g. impulse responses or variance decompositions) involving models such as structural VARs. But, in this paper, I will draw out the main ideas using the multivariate regression model:

$$y_t = x_t' \beta + \varepsilon_t \quad (1)$$

where  $y_t$  contains  $N$  dependent variables and  $x_t$  contains  $K$  explanatory variables. Note that this equation allows for many dependent variables and, if the explanatory variables are lags of the dependent variables, defines a VAR. For much of this paper, the errors,  $\varepsilon_t$ , are assumed to have a covariance matrix  $\Sigma$ . However, more flexible, heteroskedastic, error structures will also be considered.

I will organize my discussion of the challenges faced in modern empirical macroeconomics in three main sections: i) those which occur due to  $K$  being large, ii) those which occur due to  $N$  being large and iii) those which occur due to  $\beta$  and/or  $\Sigma$  changing over time. Following these three sections, I present a brief overview of a Big Data empirical application which gives a concrete example of how these challenges arise and how they can be met using Bayesian methods.

This paper is not intended to be a survey paper, providing a complete listing of all relevant citations. Nor is it intended to be a technical paper providing complete details of the econometric methodologies. I have tried to organize it around various themes, giving an intuitive description of what researchers are doing in the field and why. In this spirit, in many places where I could cite many papers, I only cite a relevant paper or two which illustrate these themes and showing the reader where to go for further study and technical details. Apologies to the many fine papers that use Bayesian methods with large macroeconomic data sets that I am not citing.

## 2 Challenge 1: Many Explanatory Variables

The macroeconomist often faces the case where many potential explanatory variables are available. Most of these variables are probably unimportant but the macroeconomist does not know which ones. A leading example of where this occurs is the cross-country growth regression literature. A common data set used by (among many others) Fernandez, Ley and Steel (2001) contains data on average per capita GDP growth for  $T = 72$  different countries along with 41 potential explanatory variables on a wide range of factors (e.g. educational attainment, investment, openness, political institutions, initial level of GDP per capita, religion, language, etc.). Data sets similar to this have been used in numerous papers which investigate the determinants of economic growth. Even with a single dependent variable, GDP growth per capita, the researcher has many coefficients to estimate with a relatively small number of observations. Cases like this abound. In time series models, where the researcher may wish to include several lags of each of several explanatory variables, the opportunities for cases where  $K$  is almost as large, or even larger, than  $T$  can also easily occur.

Why not use conventional econometric methods (e.g. non-informative prior Bayesian methods, least squares or maximum likelihood) in such a context? There is simply not enough information in the data to provide precise estimates of the parameters. Intuitively,  $T$  (sample size) reflects the amount of information in the data whereas  $K$  (the number of

explanatory variables) reflects the dimension of things the researcher is trying to estimate with that data. If  $K$  is large relative to  $T$ , the researcher is trying to do too much with too little information. This will typically lead to very imprecise inference. In the case where the number of explanatory variables is greater than the number of observations, conventional methods may simply be infeasible (e.g. in the standard formula for the ordinary least squares estimator,  $(X'X)^{-1}X'y$ , the matrix  $X'X$  will be singular and its inverse cannot be taken). If  $K < T$  a method such as least squares or maximum likelihood will produce numbers, but the Bayesian researcher will obtain large posterior variances and the frequentist will obtain wide confidence intervals. Bayesian prior information (if the researcher has it) provides more information to surmount this problem. For instance, the formula for the point estimate of  $\beta$  in the Normal linear regression model using a standard (natural conjugate) prior is  $(\underline{V}^{-1} + X'X)^{-1}(\underline{V}^{-1}\underline{\beta} + X'y)$  where  $\underline{\beta}$  and  $\underline{V}$  are the prior mean and covariance matrix. Unlike the OLS estimator, this Bayesian point estimate will exist regardless of how large  $K$  is relative to  $T$ , provided  $\underline{V}$  is a positive definite matrix. So use of Bayesian prior information can correct the problems caused by a shortage of data information. However, when  $K$  is large relative to  $T$  the role of prior information becomes increasingly important. Hence, it becomes essential to think carefully about what sensible priors are. One can interpret much of the recent Bayesian literature as developing such priors, often using hierarchical priors (a term which I will define below) which allow for simple and automatic selection of  $\underline{\beta}$  and  $\underline{V}$  (or analogous hyperparameters in more complicated models).

Related to these problems are over-parameterization and over-fitting concerns. With so many dimensions to fit, somewhere a conventional method such as OLS will fit the noise in the data rather than the pattern. A frequent consequence is apparently good in-sample fit (e.g. one might find a high  $R^2$ ), but poor forecasting performance as the spurious patterns in the noise that have been fit in-sample do not hold out-of-sample.

At this stage, the reader may be wondering as to why all the variables should be used in the first place. Why not just focus on  $k \ll K$  variables of interest and omit the rest? E.g. why should the researcher interested in the impact of education levels on economic growth not simply include the three or four variables reflecting educational attainment and ignore the rest. In the regression context, this strategy can be easily dismissed on the grounds of omitted variables bias. Using OLS methods, if important predictors for economic growth are excluded and they are correlated with education, then OLS estimates of the coefficients on the education variables will be biased. More generally, omitting important predictors for economic growth will lead to mis-specification. There are additional reasons for not focussing solely on a small number of variables of interest and discarding the rest. For instance, the forecaster typically takes the view that more information is better than less, so if more variables containing more relevant information are available, they should be used. In the structural VAR literature, where concerns about the fundamentalness of representations abound (see, e.g., Forni and Gambetti, 2014), there is often a need to include as many variables as possible in a VAR.

Another possible solution to the problems caused by a plethora of explanatory variables might be to select a more parsimonious model using hypothesis testing methods. But, this, too, runs into problems. Such an approach ignores model uncertainty since

it assumes the model selected on the basis of hypothesis tests is the “true” one which generated the data. If we have a regression with  $K$  potential explanatory variables, then there are  $2^K$  possible restricted models which include some sub-set of the  $K$  variables. In the cross-country growth regression example the  $K = 41$  potential explanatory variables imply  $2^K = 2,199,023,255,552$  possible regressions that could be run. When  $K$  is large, treating one model as if it were “true” and ignoring the huge number of remaining models is problematic. No model selection procedure is perfect, and the researcher is always uncertain about any chosen model. We want a statistical methodology that reflects this uncertainty. The fact that the selected model has been chosen using hypothesis testing procedures adds weight to the preceding criticism due to the pre-test problem. That is, conventional p-values used for deciding whether to accept or reject a hypothesis are derived assuming a single hypothesis test has been done. If a sequence of hypothesis tests is done (e.g. an initial hypothesis test suggests a variable can be omitted and then additional hypothesis tests are done on a model which omits this variable), then significance levels require adjustment. If you do one t-test using the 5% level of significance, you are correct if you take the critical value from the Student-t distribution in the standard way. If you then do a second t-test conditional on the result from a first test, this is no longer true. With  $2^K$  potential models and, thus, a huge number of possible tests, the necessary adjustments to critical values can be large and the pre-test problem can be serious in Big Data problems.

A point to highlight at this point is computation. In the cross-country growth example, with 2,199,023,255,552 possible models to consider, it may simply be computationally infeasible to estimate them all. Computational issues are important in Big Data problems and sometimes place constraints on what the researcher can do or require the need for the development of clever algorithms. The recent Bayesian literature is replete with various algorithms, usually involving simulation (either of the parameters or of the models) to get over computational hurdles.

It is also worth mentioning another problem with conventional hypothesis testing that we will return to in a subsequent section. If parameter change exists, an explanatory variable might be important at some points in time but not others. In such a case, testing whether a variable is important or not is not addressing the right question. We may not be interested in testing *whether* a variable is important, but *when* it is. For instance, a Phillips curve relationship might hold in some time periods (and, thus, a measure of the output gap is a good predictor for inflation) but not in other time periods (and, thus, a measure of the output gap would not be a good predictor in these time periods).

Thus far, we have highlighted the problems which occur with regression modelling with Big Data. It is these problems which have led many Bayesian researchers to, in the words of Hal Varian (Varian, 2014), try to develop “new tricks for econometrics”. In regression, many new methods are being developed. A popular strategy, particularly in the economic growth literature, is to use Bayesian model averaging (BMA). Instead of aiming to select a single model and presenting estimates or forecasts based on it, BMA involves taking a weighted average of estimates or forecasts from all models. The weights are the posterior model probabilities that are provided in a Bayesian analysis.<sup>3</sup> The theoretical justification

---

<sup>3</sup>For understanding the argument, it is not necessary to know precisely what posterior model prob-

for BMA can be described very simply. Let  $M_r$  for  $r = 1, \dots, R$  denote  $R$  different models and  $\phi$  be a parameter to be estimated (or a function of parameters) or a variable to be forecast. Also let  $p(M_r|Data)$  denote the posterior model probability and  $p(\phi|Data, M_r)$  be the posterior of  $\phi$  in a particular model. Then the rules of probability imply:

$$p(\phi|y) = \sum_{r=1}^R p(\phi|y, M_r) p(M_r|y). \quad (2)$$

Thus, the posterior for  $\phi$  is a weighted average of its posterior in each individual model with weights proportional to  $p(M_r|y)$ . Note that such a strategy allows for a formal treatment of model uncertainty. That is, unlike model selection procedures which choose a single model and proceed as though it were true, (2) explicitly incorporates the fact that we are only  $p(M_r|y)$  sure that  $M_r$  generated the data.

How these general ideas are operationalized depends on the particular model set-up. Moral-Benito (2015) is a recent survey paper that explains the most popular approaches (including extensions for working with panel data), provides details of several popular priors and lists some of the growing number of economic applications that use BMA methods with Big Data. In the present paper, it is only worth mentioning some of the issues that these methods address. Note first that, in the presence of  $2^K$  models and the associated computational burden, with  $K = 41$  it is going to be impossible to estimate every single model. If each model could be estimated in 0.001 seconds, it would take hundreds of years to estimate 2,199,023,255,552 models. Accordingly, simulation-based algorithms have been developed to do BMA in a computationally efficient manner (see chapter 11 of Koop, 2003). Furthermore, computation is simplified by only considering priors which lead to analytical formulae for the posterior and posterior model probabilities (in contrast to many Bayesian approaches which require the use of computationally-intensive Markov Chain Monte Carlo, MCMC, methods). Priors are also chosen so as to be automatic and objective. With so many models, the researcher cannot possibly hope to carefully choose a prior for each one. Instead, approaches such as the g-prior, which require no subjective prior input from the researcher other than the choice of a single parameter,  $g$ . This scalar controls the weight placed on prior information relative to data information.

In practice, BMA typically ensures parsimony by putting most weight on many small models. In the cross-country growth example, the regression with 41 explanatory variables will be over-parameterized. BMA will often place most of the probability on, e.g., ten or a hundred models with only a few explanatory variables. It decides, in a data-based fashion, what weight should be attached to which variables and in which models.

The basic algorithm used with BMA can also be used to do Bayesian model selection (BMS). That is, the researcher can simply choose the model with the highest posterior model probability and work with it. A drawback to this strategy is that it ignores model uncertainty. But it can be an attractive approach for the researcher who wishes to work

---

abilities or how they calculated. They are similar to information criteria. Indeed the popular Bayesian information (or Schwarz) criterion was developed as an approximation to the log of the posterior model probability.

with a single model, but does not wish to use hypothesis testing methods to select its variables.

Another class of methods for dealing with Big Data problems in regression models involves ensuring parsimony through prior shrinkage. The idea is to shrink coefficients towards something sensible (usually zero) to mitigate over-fitting. Examples include De Mol, Giannone and Reichlin (2008) and Fruhwirth-Schnatter and Wagner (2010) with Korobilis (2013b) offering a comparison of several different approaches in the context of time series regression models. Recently, the LASSO (see Park and Casella, 2008) and a range of similar prior shrinkage algorithms with names like stochastic search variable selection (SSVS) and spike and slab priors have become popular.

I will use SSVS to demonstrate the basic ideas underlying this kind of approach. But first, I will briefly describe the general idea of how Bayesian priors can be used to do shrinkage. Suppose you have a Normally distributed prior for a single regression coefficient,  $\beta$ :

$$\beta \sim N(\underline{\beta}, \underline{V}).$$

Bayesian analysis involves combining prior with data information. With a natural conjugate prior, the Bayesian posterior estimate will be a combination of the OLS estimate and the prior mean,  $\underline{\beta}$ . Thus, it can be interpreted as taking a data based quantity (such as OLS) and shrinking it towards  $\underline{\beta}$ . The prior variance,  $\underline{V}$ , controls the degree of shrinkage since it reflects the confidence the researcher has that  $\beta$  is near  $\underline{\beta}$ . To put it another way, to the question: “towards what should the coefficient be shrunk?” the answer is  $\underline{\beta}$ . To the question: “by how much should the coefficient be shrunk?” the answer is controlled by  $\underline{V}$ . Small values of  $\underline{V}$  are consistent with strong shrinkage and large values with little shrinkage. In the limit, as  $\underline{V} \rightarrow \infty$ , the prior becomes non-informative and no shrinkage is done at all (i.e. the OLS estimate is obtained). In a conventional, subjective, Bayesian analysis  $\underline{\beta}$  and  $\underline{V}$  are selected by the researcher to reflect prior beliefs about likely values for  $\beta$ .

The modern methods used with Big Data regressions such as SSVS do shrinkage in a more objective fashion. By this I mean that the data decides how much shrinkage should be introduced in the prior for each coefficient. The SSVS prior for  $\beta$  can be written as:

$$(1 - \gamma) N(0, \tau_0^2) + \gamma N(0, \tau_1^2)$$

where  $\gamma = 0$  or  $1$  and  $\tau_0$  is small and  $\tau_1$  is large. It can be seen that, if  $\gamma = 0$ ,  $\beta$  is shrunk towards zero whereas if  $\gamma = 1$ ,  $\beta$  is freely estimated from the data with little shrinkage.  $\gamma$  is treated as an unknown parameter which is estimated from the data. Thus, the prior reflects only a belief that it is likely many of the potential explanatory variables are likely to be unimportant, but it does not reflect any belief about which ones these are. The data is used to decide which coefficients should be shrunk to be very close to zero (or, in the limit, to be shrunk to zero and thus deleting the variable from the regression model) and which variables should be freely estimated. This is an example of a hierarchical prior which is one where the prior for a parameter (here  $\beta$ ) depends on yet another parameter ( $\gamma$ ) which in turn requires a prior. Most of the Bayesian shrinkage methods (including



the LASSO) can be expressed as hierarchical priors with similar motivation as the SSVS prior.

This section has provided a small sample of the many Bayesian econometric methods which are used successfully with Big Data regressions. But regressions, involving only a single dependent variable, are only one tool used by the macroeconomist. Multiple equation models such as VARs are even more popular and it is with them that the Big Data issues become even more important.

### 3 Challenge 2: Many Dependent Variables

There has been an explosion of recent work in many fields of macroeconomics involving large VARs. In terms of the notation given in (1), this means  $y_t$  is a vector of  $N$  dependent variables. The VAR is obtained by suitably formatting  $x_t$  so as to contain lags of the dependent variables, deterministic terms and exogenous variables. For many years, VARs have been extensively used for small values of  $N$  (e.g.  $N$  between 2 and 10). But recently researchers have begun working with VARs where  $N$  is large. Values for  $N$  of about 20 are quite common, but ambitious researchers are pushing into the region where  $N > 100$ . A pioneering paper in this literature is Banbura, Giannone and Reichlin (2010) which uses a standard set of over 100 US macroeconomic variables similar to those now easily available at FRED-MD (see McCracken and Ng, 2015). Previously macroeconomists working with data sets of this magnitude had used factor-based methods. Why did the Banbura, Giannone and Reichlin (2010) paper attract a great deal of attention? Put briefly, it showed large VAR methods worked. They tended to forecast better than factor methods. Furthermore, impulse responses were sensible and the large VAR revealed structural insights different from smaller VARs. Thus, two of major activities of the empirical macroeconomist, forecasting and structural analysis, benefitted from working with Big Data in a new way.

In Banbura, Giannone and Reichlin (2010) the Big Data aspect arose since the VAR involved many US macroeconomic variables. There are also large VAR applications using similar macroeconomic data sets for other countries. For instance, Bloor and Matheson (2010) work successfully with a large VAR using New Zealand data. There are also applications where the Big Data aspect arises since data for many countries is used. For instance, Carriero, Kapetanios and Marcellino (2010) use a large VAR to model exchange rates for many countries and Koop and Korobilis (2016) investigate the eurozone sovereign debt crisis using countries in the eurozone. In Carriero, Kapetanios and Marcellino (2012), Big Data occurs from a desire to model US government bond yields of different maturities. The wish to forecast the components of euro area inflation leads to the large VAR used in Giannone, Lenza, Momferatou and Onorante (2014). Large VARs have also been used for reasons other than pure forecasting or structural analysis in papers such as Bańbura, Giannone and Lenza (2015) and Jarociński and Maćkowiak (2016). The former of these does conditional forecasting and scenario analysis using a euro area data set whereas the latter develops methods for investigating Granger-causality and Granger-causal-priority. This brief and very incomplete listing of large VAR empirical papers is intended to illustrate the range of applications in which they are being found useful.

The wish to work with large VARs is based on their empirical success and a desire to include all available information. But the problems of overparameterization discussed above in the context of the single-equation regression model are greatly magnified in the multiple equation VAR. Each equation of the VAR contains lags of all of the dependent variables as well as any deterministic terms and exogenous variables. Even if we consider only the lagged dependent variables, the number of VAR coefficients can be huge. A large VAR with quarterly data might have  $N = 100$  variables and  $p = 4$  lags which leads to 40,000 coefficients. With monthly data, researchers often use longer lag lengths and thus a large VAR with monthly variables can have over 100,000 coefficients. The error covariance matrix,  $\Sigma$ , is also parameter rich since it contains  $\frac{N(N+1)}{2}$  distinct parameters which is 5,050 for the  $N = 100$  example.

The large VAR world is, thus, one where the number of parameters may far exceed the number of observations. As we have seen, in theory, this is no problem for Bayesian methods. These combine likelihood function with prior. Even if the parameters in the likelihood function are not identified, combining the likelihood with the prior will (under weak conditions) lead to valid posterior inference. However, the role of prior information becomes more important as the likelihood becomes less informative relative to the prior. In the large VAR literature, the Minnesota prior and some of its close relations are most commonly used. The Minnesota prior grew out of work done at the University of Minnesota and the Federal Reserve Bank of Minneapolis (see Doan, Litterman and Sims, 1984). Technical details about the Minnesota prior can be found in many places, including Koop and Korobilis (2010), Blake and Mumtaz (2012) and Karlsson (2013). Two points are worth noting here. First, Bayesian analysis using the Minnesota prior is computationally simple, involving only analytical results and with no need to use MCMC methods. This has made it ideal for scaling up to large VARs where other approaches which require MCMC methods are computationally infeasible. Second, it is a shrinkage prior where coefficients are pulled towards a set of prior beliefs that reflect the empirical wisdom of its developers. These beliefs include the ideas that own lags (e.g. lags of GDP in the equation where GDP is the dependent variable) are more important than other lags (e.g. lags of GDP in the inflation equation); that more distant lags are less important than more recent lags; and that the first own lag is likely to be important in each equation if the researcher is working with levels data (i.e. it shrinks towards individual random walks). These properties make the Minnesota prior popular and increasingly, computer software such as the BEAR toolbox of Dieppe, Legrand and van Roye (2016) make it simple to estimate large VARs with many different forms of the Minnesota prior or related variants.

But the Minnesota prior and related priors are subjective in the sense that the decisions about which variables to shrink in which manner are made by the researcher. Although papers such as Giannone, Lenza and Primiceri (2015) are developing methods for estimating the degree of shrinkage, the general nature of this shrinkage (e.g. that other lags are shrunk less than own lags) is chosen by the researcher. In the preceding section, methods such as SSVS and the LASSO were discussed. These involved hierarchical priors which were less subjective than conventional non-hierarchical priors in the sense that they only expressed a belief that many coefficients were probably unimportant and could be shrunk to zero. These hierarchical priors are agnostic about which ones should be shrunk. This kind of prior information can also be used with large VARs. For

instance, SSVS methods were first introduced to the VAR literature by George, Sun and Ni (2008) but have since been used successfully with larger VARs in Koop (2013) and Korobilis (2013a). It is worth noting that SSVS methods can also be used on the error covariance matrix,  $\Sigma$ , thus helping ensure parsimony in the large VAR. Korobilis (2013a) also considers the use of LASSO methods in large VARs and these are further developed in Gefang (2014). At present, these approaches have only been used with VARs containing fewer than 50 dependent variables. Their need for computationally intensive MCMC methods has precluded their use in really large VARs with  $N = 100$  or more. It will be interesting to see if these methods, which have worked so well with fairly large VARs, can be scaled up to the extremely large VARs in use today.

One way that large VARs can emerge is through the use of data for multiple countries. If the macroeconomist has, say, 10 variables for each of 20 countries and wants to jointly model them in a VAR, this leads to  $N = 200$ . In this regard, it is worth noting that priors adapted for the multi-country nature of the data have been developed. That is, instead of simply using a Minnesota prior or SSVS prior for the large VAR, priors developed for the panel structure of the data exist. Canova and Ciccarelli (2013) surveys the panel VAR literature. Koop and Korobilis (2016) adapt and extend earlier approaches to deal with a large panel VAR. In the final section of this paper, Bayesian estimation of a time-varying parameter panel VAR will be discussed as an illustration of the issues involved.

Another category of approaches to large VAR estimation can be interpreted as compressing the data to produce a smaller VAR which is computationally manageable. To illustrate the basic ideas of these approaches, consider the VAR model with one lag and no deterministic terms or exogenous variables:

$$y_t = Ay_{t-1} + \varepsilon_t, \quad (3)$$

where all assumptions are as in (1). The matrix of VAR coefficients,  $A$ , is  $N \times N$  and, thus, there are a huge number of free parameters to be estimated in the large VAR. This number gets proportionally larger with longer lag lengths. There are several recent approaches that re-write (3) as

$$y_t = B(\Phi y_{t-1}) + \varepsilon_t$$

where  $B$  is  $N \times m$  and  $\Phi$  is  $m \times N$  where  $m \ll N$ .  $\Phi Y_{t-1}$  can be interpreted as compressed data and there are various methods of doing this compression. Once the compression is done the macroeconomist can work with a much smaller model. Models of this form have gone by various names (e.g. reduced rank model, vector error correction model) and used in various contexts in macroeconomics. Under the name multivariate autoregressive index model, they are being successfully used in large VARs by Carriero, Kapetanios and Marcellino (2015). These authors interpret  $\Phi$  as an unknown matrix of parameters to be estimated and develop MCMC methods for doing so. This looks like a very interesting avenue for working with large VARs, but it does still require computationally burdensome MCMC methods that may not be feasible when  $N = 100$  or more.

Compressed data methods are commonly used in fields such as machine learning and image recognition as a way of projecting information in data sets with a huge number of variables into a much lower dimension before analysis. Theoretical results show that,

under weak conditions, the information lost through such compression may be small, but the computational gains may be huge. A lesson of this literature is that compression can be done randomly using simple schemes, without reference for the data. Such so-called data-oblivious schemes are computationally very efficient since the researcher only needs to work with the entire (huge) data set once (i.e. when the compression is being done and  $\Phi y_{t-1}$  is produced). Subsequently statistical analysis can be done only using this compressed data. These features are exploited in Koop, Korobilis and Pettenuzzo (2016) who develop random compression methods for the VAR and demonstrate their successful forecast performance.

In the preceding discussion of Big Data methods for regression, BMA was mentioned. Instead of selecting a single regression, the researcher could average over all regressions with weights proportional to their posterior model probabilities. This often leads to the researcher placing most weight on many small parsimonious models, thus surmounting some of the problems caused by working with regressions where the number of explanatory variables is huge relative to the sample size. Similar schemes are used with large VARs, but in a time-varying fashion leading to Dynamic Model Averaging (DMA). If the methods are used for model selection then Dynamic Model Selection (DMS) results. These methods are mostly used with forecasting (although they can be used for structural analysis as well). The idea underlying DMS is that the forecasting model can change over time. The idea underlying DMA is that the weights used in model averaging can change over time.

We explain the general ideas underlying DMA and DMS where  $M_r$  for  $r = 1, \dots, R$  are a set of models under consideration and  $\mathcal{D}_{t-1}$  is the data available at time  $t - 1$ . Interest centres on forecasting  $y_t$  given  $\mathcal{D}_{t-1}$ . The key quantity of interest is  $p(M_r|\mathcal{D}_{t-1})$  which is the probability attached to forecasts from  $M_r$ . DMA takes forecasts from all models and averages using these probabilities. DMS forecasts using the model with the highest  $p(M_r|\mathcal{D}_{t-1})$ . A problem with a formal Bayesian DMA or DMS approach is computation. There are  $2^{TR}$  possible paths of models over time if we consider every possible combination of model and time period. This is too large unless  $R$  and  $T$  are both small. Raftery, Karny and Ettler (Tech, 2010) surmount this problem through the use of an approximation. They use so-called forgetting factor methods which allow for the calculation of  $p(M_r|\mathcal{D}_{t-1})$  in a fast, recursive manner, in the spirit of Kalman filtering. Since this pioneering paper, DMA methods have been used in many macroeconomic and financial applications and are just beginning to be used in Big Data problems.

Bayesians define a model as involving a likelihood function and a prior. The models used in DMA can be any consistent with this definition. Raftery, Karny and Ettler (2010) considered a set of regression models which differed in their explanatory variables. But DMA can be used with models which differ in other aspects (including having different priors). In the VAR context, Koop (2014) uses DMA methods with VARs of different dimensions. Thus, the algorithm can choose to forecast with the large VAR at some points in time and smaller more parsimonious VARs at other times. The decision as to which dimension to choose is made based on recent forecast performance.

## 4 Challenge 3: Parameter Change

There is evidence that the properties of most macroeconomic variables change over time (see, among many others, Stock and Watson, 1996). When these variables are put into an econometric model, it becomes important to allow for parameter change to accommodate this. This parameter change can be of various sorts. To illustrate this point, consider the discussions surrounding U.S. monetary policy and, in particular, the question of whether the high inflation and high unemployment rate of the 1970s were due to bad policy or bad luck. Some have argued that the way the Federal Reserve reacted to inflation has changed over time. After 1980, the Federal Reserve became more aggressive in fighting inflation pressures than before. This is the bad policy story. In the 1970s the Federal Reserve was pursuing inappropriate policies which were subsequently changed. Any econometric model should incorporate such a change in the monetary policy transmission mechanism. In terms of a multivariate time series model such as a VAR, this means the VAR coefficients change over time (e.g. some of the VAR coefficients will control how past values inflation impact on the interest rate decisions of the Federal Reserve and these will have been different in the 1970s than subsequently).

However, there are other researchers who argue that the stagflation of the 1970s was due to exogenous shocks (e.g. the OPEC oil price shock) that the Federal Reserve had no control over. This is the bad luck story and it is fundamentally one involving error variances. It says that the 1970s were a highly volatile time which was succeeded by the Great Moderation of the business cycle in the 1980s and 1990s. If this is the case, then the error covariance matrix,  $\Sigma$ , in the VAR must become  $\Sigma_t$  and be time-varying.

The bad policy versus bad luck debate illustrates how, in order to shed light on an important macroeconomic issue using a VAR, it is essential to allow for both VAR coefficients and the error covariance matrix to change over time. Many other macroeconomic questions of interest involve similar considerations. Proper treatment of time-variation in parameters can be difficult, even with small data sets, but with Big Data the problems are magnified. Nevertheless, some promising methods are being developed as will be discussed in this section.

There are many different approaches used to model time-variation in parameters. For instance, various models allow for regime change or structural breaks where the coefficients change abruptly according to the state of the business cycle or at a point in time. These include Markov switching models (see, e.g., Sims and Zha, 2006) and threshold or smooth transition threshold models of various sorts (e.g. Koop and Potter, 2006). But much recent work has used time-varying parameter (TVP) models which allow for gradual change in parameters. In terms of (1), these models replace the assumption that  $\beta$  is constant by

$$\beta_t = \beta_{t-1} + u_t.$$

In words, the researcher's best guess about what the current VAR coefficients are is that they will be the same last period's. However, there is some unpredictable error causing change over time. Examples of empirical macroeconomic papers which use TVP-VARs are too numerous to adequately cite here. A few prominent early citations are Cogley and Sargent (2001, 2005) and Primiceri (2005). TVP-VARs are particularly popular since

they are state space models for which standard state space methods involving the Kalman filter and state smoother can be used.

The preceding discussion was in terms of time-variation in VAR coefficients. However, allowing for time variation in error covariance matrices is also important and may even be more important in many applications. Indeed the best fitting model of Sims and Zha (2006) involves nine different regimes all of which differ only in the error covariance matrix. And the main difference in econometric specifications between Cogley and Sargent (2001) and Cogley and Sargent (2005) is that the former is homoskedastic while the latter is heteroskedastic, reflecting the increasing realization of the importance of allowing for volatilities to change over time.

The papers I have just cited, as well as much of the related literature, is Bayesian. This reflects a desire to use prior shrinkage as a way of mitigating the over-parameterization concerns. That is, over-parameterization concerns already arise in constant-coefficient VAR models. Allowing for the VAR coefficients and/or error covariance matrices to change means more parameters to estimate and even more concerns about over-parameterization. It is also important to note that dealing with all these challenges raises the computational burden, sometimes massively. That is, with the constant coefficient VAR with Minnesota prior analytical formulae are available for the posterior and one-step ahead predictive densities. However, as soon as we leave the constant-coefficient homoskedastic world, MCMC methods are typically required. These can be computationally daunting even in the context of estimating a small model. But in larger models or when carrying out a recursive pseudo out-of-sample forecasting exercise, which would require repeatedly running the MCMC algorithm on an expanding window of data (see, e.g., D’Agostino, Gambetti and Giannone, 2013), the computational burden can become prohibitive.

I stress that all the statements and citations so far given in this section do not relate to Big Data. The applications cited involve TVP-VARs or Markov switching VARs with a small number of variables. For instance, the TVP-VAR with multivariate stochastic volatility of Primiceri (2005) has three dependent variables. The development of Bayesian methods which allow for time-variation in parameters in large VARs is in its infancy. The concerns regarding over-parameterization and computation, present with small VARs, become huge with large VARs. Koop and Korobilis (2013) develop methods for large TVP-VARs but these involved approximations, thus avoiding the need for MCMC methods but at the cost of not producing the exact posterior and predictive densities for the large TVP-VAR. Koop, Korobilis and Pettenuzzo (2016) adopt similar approximations when extending their compressed VAR approach to allow for time-varying parameters and error covariances. Very recently, Kapetanios, Marcellino and Venditti (2016) have developed an interesting way of estimating large TVP-VARs which, although not fully Bayesian, can be given a quasi-Bayesian interpretation. However, in general, there is a need for a fully Bayesian analysis of large TVP-VARs.

In terms of exact Bayesian posterior analysis with extensions of large VARs, most progress has been made with VARs where the VAR coefficients are constant over time, but the error covariance matrix is allowed to change. This reflects the common empirical finding that with many macroeconomic data sets, there is more parameter change in error variances than in the VAR coefficients. Without getting into too much complicated econometric theory, the basic idea in this new line of research can be explained by noting

that the VAR error covariance matrix for the entire vector of  $TN$  errors can be written as  $\Sigma \otimes I$  in the homoskedastic VAR where  $\Sigma$  is an  $N \times N$  matrix allowing for the errors in the different equations to be correlated with one another and  $I$  is the  $T \times T$  identity matrix embodying the assumption that the errors are uncorrelated over time. This Kronecker structure, along with a similar structure in the prior, is crucial for keeping computation manageable in large VARs. If we let  $\beta$  be the vector of coefficients in the VAR(p) and  $K$  be the number of right hand side variables in each equation ( $K = Np$  in a VAR without deterministic terms or exogenous variables), then the posterior covariance matrix for  $\beta$  is an  $NK \times NK$  matrix. Working with it involves lots of matrix manipulations such as matrix inversion. Inverting an  $NK \times NK$  matrix can be extremely difficult for the computer when, say,  $N = 100$  and  $K = 1200$  and, thus, a  $120000 \times 120000$  matrix must be inverted. But the Kronecker structure (roughly speaking) turns this into a problem of inverting an  $N \times N$  matrix and a  $K \times K$  matrix separately which is much more feasible in large VARs. To give the reader an idea of the order of magnitude of the computational savings, here is a quotation from a recent paper in relation to the posterior covariance matrix of the VAR coefficients: “in a system of 20 variables and 4 lags, the manipulation of this matrix would involve 4,251,528,000 elementary operations, but when the matrix has a Kronecker structure, the number of operations reduces to just 539,441” (Carriero, Clark and Marcellino, 2015, page 8).

This basic idea – that we must keep a particular Kronecker structure (or something close to it) in the error covariance matrix and prior in order to keep computation feasible – has influenced several recent papers and allows the researcher to work with heteroskedastic large VARs using fully Bayesian methods. Carriero, Clark and Marcellino (2015) develop a model of common drifting volatility in large Bayesian VARs using the assumption that

$$\Sigma_t = f_t \Sigma$$

where  $f_t$  is a univariate stochastic volatility process. In words, there is one common factor driving volatilities of all variables. In many applications, this assumption may be reasonable, but in others this may be too restrictive. Subsequently, Carriero, Clark and Marcellino (2016) develop Bayesian methods for estimating a more flexible version of this model that maintains the common drifting volatility process but also allows for the error variance in each equation to follow a different stochastic volatility process.

Another recent paper worth noting is Chan (2015) who develops efficient computational methods for large VARs with error covariances having the form  $\Sigma \otimes \Omega$  for general  $\Omega$ . The  $T \times T$  matrix  $\Omega$  allows for different types of error correlation over time. Different choices for  $\Omega$  allow for different extensions including common stochastic volatility, moving average structures of certain types and fat-tailed errors. Since any or all of these may be present in the data and, thus, may be important to model, the ability to do fully Bayesian work with error covariances of the  $\Sigma \otimes \Omega$  form is a key step forward.

Approaches such as Carriero, Clark and Marcellino (2015, 2016) and Chan (2015) have been shown to work well in fairly large VARs (e.g. VARs involving 20 – 50 variables) but not the 100 or more that are increasingly used in macroeconomic research. And, unlike papers such as Primiceri (2005) who use extremely flexible multivariate stochastic volatility processes for  $\Sigma_t$ , they use specifications for  $\Sigma_t$  which may be restrictive in some

applications. Hence, there is much work to be done to develop Bayesian methods for large VARs with time-varying parameters. But, for homoskedastic VARs, the tools are there. And methods for allowing the time-variation in parameters (both in VAR coefficients and in the error covariance matrix), tools exist for an analysis of fairly rich structures and more are being developed all the time. We are at a stage where the empirical macroeconomist working with Big Data can work with some very useful (and empirically necessary) Bayesian tools even with time-varying parameter models.

## 5 A Large Time-Varying Panel VAR

This paper has discussed three important challenges facing the empirical macroeconomist in the presence of Big Data and how Bayesian methods are increasingly used to address them. A major theme that has emerged in this discussion is that, since more information is often better, it is potentially important to work with Big Data. However, much of this information may be irrelevant or just be noise and, thus, methods for ensuring parsimony such as prior shrinkage are necessary. A second theme is that conventional statistical methods, designed for Small Data, may not work well with Big Data and that Bayesian methods offer a plausible and increasingly practically feasible alternative. The discussion was kept fairly general and non-technical. In this section, I will discuss a modelling strategy used in Koop and Korobilis (2015) as a way of showing how these themes infuse my work in practice. The hope is that the general ideas discussed previously will become clearer in a practical context.

In an increasingly globalized world, macroeconomists are increasingly interested in jointly modelling data from several countries. That is, interlinkages between countries can be important in many macroeconomic and financial contexts (e.g. the eurozone crisis) and they can only be modelled using data from many countries. The wish to work with data for many countries is a prime example of how Big Data can arise. Even if you have a relatively small set of variables for an individual country (e.g. wish to work with 5-10 variable in a VAR), when working with many countries (e.g. the 19 countries in the eurozone), the researcher can easily end up with 100 or more variables of interest. In the multi-country VAR literature, problems caused with this profusion of variables are often addressed by placing restrictions on the way countries interact. For instance, the global VAR literature (see, e.g., Dees, Di Mauro, Pesaran and Smith, 2007) imposes restrictions that data from foreign countries can only affect a domestic economy through global averages. But what if global averages are not the correct specification? For instance, if a small country like Greece had a disproportionate effect on other countries at the time of the eurozone crisis this could be missed if subsumed in a global average. In such a case, the researcher may wish to work with an unrestricted model such as a VAR for all the variables. Such a VAR will be very large and, hence, the need for methods such as those discussed in this paper arise.

Koop and Korobilis (2015), in the context of a eurozone inflation forecasting exercise, work with  $G = 19$  eurozone countries and  $N = 7$  variables. If jointly modelled in an unrestricted VAR this would mean a large 133 variate VAR. Their data set runs from 1999M1-2014M12. Hence, this is a case where the data information is small relative to



the number of parameters being estimated. Parsimony is achieved through a combination of various methods described earlier in this paper.

Relative to an completely unrestricted multi-country VAR, the first way parsimony can be induced is through the assumption that the huge number of VAR coefficients are not completely unrelated to one another but reflect some underlying structure due to the nature of the problem. An influential early paper working with multi-country VARs is Canova and Ciccarelli (2009). If  $\beta$  is the vector of VAR coefficients, then they suggest a reasonable structure for it in a multi-country context is:

$$\begin{aligned}\beta &= \Xi_1\theta_1 + \Xi_2\theta_2 + .. + \Xi_q\theta_q + e \\ &= \Xi\theta + e,\end{aligned}\tag{4}$$

where  $\Xi = (\Xi_1, ..., \Xi_q)$  are known matrices and  $\theta = (\theta'_1, ..., \theta'_q)'$  is an  $R \times 1$  vector of unknown parameters with  $R$  being small. In their application, they define  $\Xi$  so as to imply the VAR coefficients are composed of a common factor, a factor specific to each country and a factor specific to each variable. That is,  $\Xi_1$  is a vector of ones and  $\theta_1$  is a scalar (this defines the factor common to all coefficients).  $\Xi_2$  is a matrix containing zeros and ones defined so as to pick out coefficients for each variable and  $\theta_2$  is an  $N \times 1$  vector.  $\Xi_3$  is defined to pick out coefficients for each country and  $\theta_3$  is an  $G \times 1$  vector.  $e$  is an error which picks up remaining heterogeneity in coefficients. Note that (in the  $p = 1$  case), instead of having to estimate the  $N^2 \times G^2$  individual coefficients of the unrestricted VAR, the researcher only needs to estimate the  $N + G + 1$  coefficients in  $\theta$  along with the parameters in the distribution for  $e$ . A point worth noting is that this is an example of a prior designed to take into account the multi-country nature of the model. The popular Minnesota prior would not reflect this and, although it may be appropriate for a conventional macroeconomic VAR for a single country, it may not be appropriate here.<sup>4</sup> In general, with Big Data prior information can be very important and it is important to have a prior that is appropriate for the data set at hand.

Adapting (4) to allow for time-varying VAR coefficients, thus producing a time-varying parameter panel VAR (TVP-PVAR) is done by Canova and Ciccarelli (2009) by allowing  $\beta_t$  to vary over time according to:

$$\begin{aligned}\beta_t &= \Xi\theta_t + e_t \\ \theta_t &= \theta_{t-1} + u_t.\end{aligned}\tag{5}$$

Thus, all the VAR coefficients can change over time, but this change is driven by a lower dimensional vector  $\theta_t$  which contains only the common factor, the factor specific to each country and a factor specific to each variable. Thus, the extremely parameter-rich TVP-PVAR is made much more parsimonious through a hierarchical prior.

Koop and Korobilis (2015) adopt the structure given in (4) and (5) and consider the common/country/variable component division suggested by Canova and Ciccarelli (2009).

---

<sup>4</sup>The BEAR software, reflecting this consideration, has a separate set of commands for the panel VAR.

Such a division sounds, in theory, sensible and in practice seems to work well in many applications. But what if it does not? One can imagine applications where cross-country heterogeneity is large and the idea of having a common factor is unreasonable. Or the assumption that a single factor for each country is enough to model such heterogeneity is insufficient. Or many other sorts of reasons why the assumptions embedded in (4) and (5) might be too restrictive. Hence, Koop and Korobilis (2015) consider more than one choice for  $\Xi$ . This strategy is consistent with a theme emphasized previously: there are good theoretical reasons for thinking Bayesian model averaging methods (either using BMA or DMA) to be an attractive thing to do and in practice averaging over many simple models is a promising way to proceed in a Big Data world. That is, by working only with small models the benefit of parsimony is gained, but by considering many of them with different variables (or, in this case, different priors) data information is not excluded.

There is much evidence that working with large VARs, involving a large number of variables, is a sensible thing to do. However, the question of whether large VARs really forecast better than small VARs is, in the end, an empirical one. Large VARs may be better than small VARs in some applications, but not others. There may be some periods where small VARs are good, other periods where large VARs win. In the case of Koop and Korobilis (2015), we might expect small VARs to do well. Forecasting inflation is hard. Often is it hard to beat simple forecasting models (see, e.g., Stock and Watson, 2007 or Faust and Wright, 2013). Hence, Koop and Korobilis (2015) acknowledge that it is possible that including  $G = 7$  variables for each country into the TVP-PVAR and/or allowing for variables in one country to influence inflation in a different country is unnecessary. Perhaps simple forecasting models are adequate. Koop and Korobilis (2015) did not wish immediately go to a smaller, parsimonious model, but discover, in a data-based fashion whether simpler models sufficed. In this spirit, they chose a set of  $G_C$  core variables of interest (inflation, unemployment rate and industrial production) and worked with VARs of dimension  $G_C$  or larger.

They also reasoned that perhaps allowing for time-variation in parameters may be unnecessary but again, wanted to establish whether constant coefficient models were adequate in a data based fashion. Accordingly they considered models with and without time-variation in VAR coefficients. They also considered different specifications for the time-varying error covariance matrix.

Koop and Korobilis (2015) thus ended up with a large set of models which differed in the choice of  $\Xi$ , the choice of VAR dimension, the choice of degree of time-variation in  $\beta$  and  $\Sigma$  as well as some other specification choices. Faced with a large number of models with up to potentially 133 variables, use of MCMC methods was computationally infeasible and, instead, they reverted to approximate methods. They used DMA and DMS methods to navigate their way through the huge range of models in a dynamic fashion. Often they found small VARs to forecast well, but there were times when the large VARs forecast better. At times the structure for  $\Xi$  suggested by Canova and Ciccarelli (2009) was chosen by DMS, but there were also times it did not. Overall, the forecasting performance of the TVP-PVAR approach of Koop and Korobilis (2015) was as good or better than other competitors and it achieved this success by switching dynamically between different specifications.

## 6 Summary and Conclusions

In this paper, I have discussed how Big Data often arises in macroeconomics, why macro-economists should often be interested in using it and the problems it poses for conventional econometric methods. I have discussed how Bayesian methods can be used to overcome these problems and given an illustration of how they can work in practice. Bayesians are not the only ones developing new methods to deal with these issues, but Bayesian methods have two features which make them attractive in this context. First, they allow for the use of prior information. Researchers in this field, regardless of whether they are Bayesian or frequentist, acknowledge the need for shrinkage methods so as to mitigate the effects of over-parameterization and over-fitting that occur with noisy macroeconomic data. The Bayesian prior provides a formal route for the researcher to introduce such shrinkage. Hierarchical priors, such as those used by the SSVS and LASSO methods discussed in this paper, allow many key shrinkage choices to be made in a data-based fashion. Loosely speaking, they allow the researcher to express a prior belief that many or most coefficients are likely zero without expressing any beliefs about precisely which ones are zero. The data is used to make the latter choice.

Second, Bayesians treat models as random variables and it is thus meaningful to attach probabilities to each. This makes it easy to work with many models and average over them – and even to do so in a time-varying fashion. With Big Data, the ability to work with numerous parsimonious models is often an attractive thing to do. Instead of selecting a single small model involving a tiny sub-set of all the potential variables, it is possible to have different models involving different sub-sets of the variables. For instance, the researcher may not want to work with all of the more than 100 variables available in the FRED-MD data base in a single model. But it is possible to work with many smaller models: each of the more than 100 variables can appear in one or more of these models and the econometric methodology can be used to decide which ones receive the most weight.

I have organized the discussion around three main challenges: those which occur in single equation models such as regression models with many explanatory variables, those which occur in multiple-equation models such as VARs and those which occur when time-variation in parameters is allowed for. Of these three challenges, the first is the one we are closest to meeting. There are a range of methods, some discussed in this paper, which have been found to work well in regressions where the number of explanatory variables is large relative to (or even larger than) sample size. Great progress has been made with the second challenge, with large VAR methods enjoying great success in a variety of applications. However, there are still many interesting issues to address with large VARs. These arise partly because the large VAR has a commensurately large error covariance matrix and partly due to the computational challenges raised by the sheer magnitude of the number of parameters or models involved. But it is in the third challenge that most needs to be done. There is little doubt that, in many applications, VAR coefficients and/or error covariance matrices are changing over time. But faced with the enormous range of potential ways these parameters might change and the over-parameterization and computational problems that must be addressed, there is much to do. But rising to this challenge will be necessary if macroeconomists are to benefit from the rich potential

of Big Data.

## References

- Bañbura, M., Giannone, D. and Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting* 31, 739-756.
- Banbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, 25, 71-92.
- Blake, A. and Mumtaz, H. (2012). Applied Bayesian econometrics for central bankers. Centre for Central Banking Studies, Bank of England, Technical Handbook – No. 4.
- Bloor, C. and Matheson, T. (2010). Analysing shock transmission in a data-rich environment: a large BVAR for New Zealand. *Empirical Economics* 39, 537-558.
- Canova, F., Ciccarelli, M. (2009). Estimating multicountry VAR models. *International Economic Review* 50, 929-959
- Canova, F. and Ciccarelli, M. (2013). Panel vector autoregressive models: A survey. *Advances in Econometrics* 32, 205-246.
- Carriero, A., Clark, T. and Marcellino, M. (2015). Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2015.1040116.
- Carriero, A., Clark, T. and Marcellino, M. (2016). Measuring uncertainty and its impact on the economy. Manuscript.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2010). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25, 400-417.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2012). Forecasting government bond yields with large Bayesian VARs. *Journal of Banking and Finance*, 36, 2026-204.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2015). Structural analysis with multivariate autoregressive index models. CEPR Discussion Paper No. DP10801.
- Chan, J. (2015). Large Bayesian VARs: A flexible Kronecker error covariance structure, <http://people.anu.edu.au/joshua.chan/papers/BVAR.pdf>.
- Cogley, T. and Sargent, T. (2001). Evolving post-World War II inflation dynamics. *NBER Macroeconomic Annual* 16, 331-373.
- Cogley, T. and Sargent, T. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics* 8, 262-302.
- D’Agostino, A., Gambetti, L. and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28, 82-101.
- De Mol, C., Giannone, D. and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* 14, 318-328.
- Dees, S., Di Mauro, F., Pesaran, M.H. and Smith, V. (2007). Exploring the international linkages of the Euro area: A global VAR analysis. *Journal of Applied Econometrics* 22, 1-38.
- Dieppe, A., Legrand, R. and van Roye, B. (2016). The BEAR toolbox. European Central Bank working paper, no. 1934.
- Doan, T., Litterman, R. and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3, 1-144.
- Faust, J. and Wright, J. (2013). Forecasting inflation. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting*, volume 2. Amsterdam: North Holland.

- Fernandez, C., E. Ley and Steel, M. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16, 53-76.
- Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and non-Gaussian state space models. *Journal of Econometrics* 154, 85-100.
- Forni, M. and Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics* 66, 124-136.
- Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting*, 30, 1-11.
- George, E., Sun, D. and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* 142, 553-580.
- Giannone, D., Lenza, M. and Primiceri, G. (2015). Prior selection for vector autoregressions, *Review of Economics and Statistics*, 97, 436-451.
- Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2014). Short-term inflation projections: a Bayesian vector autoregressive approach, *International Journal of Forecasting*, 30, 635-644.
- Hendry, D. and Doornik, J. (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge: MIT Press.
- Jarociński, M. and Maćkowiak, B. (2016). Granger-causal-priority and choice of variables in vector autoregressions. *Review of Economics and Statistics* doi:10.1162/REST\_a\_00601.
- Kapetanios, G., Marcellino, M. and Venditti, F. (2016). Large time-varying Parameter VARs: A non-parametric approach. Manuscript.
- Karlsson, S. (2013). Forecasting with Bayesian vector autoregressions. Pages 197-897 in Graham Elliott and Allan Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2, Elsevier.
- Koop, G. (2003). *Bayesian Econometrics*. Chichester: Wiley.
- Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* 28, 177-203.
- Koop, G. (2014). Forecasting with dimension switching VARs. *International Journal of Forecasting* 30, 280-290.
- Koop, G. and Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics* 3, 267-358.
- Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics* 177, 185-198.
- Koop, G. and Korobilis, D. (2015). Forecasting with high dimensional panel VARs. Manuscript available at <https://sites.google.com/site/garykoop/research>
- Koop, G. and Korobilis, D. (2016). Model uncertainty in panel vector autoregressive models. *European Economic Review* 81, 115-131.
- Koop, G., Korobilis, D. and Pettenuzzo, D. (2016). Bayesian compressed vector autoregressions. Manuscript available at <https://sites.google.com/site/garykoop/research>.
- Koop, G. and Potter, S. (2006). The vector floor and ceiling model. Chapter 4 in *Nonlinear Time Series Analysis of the Business Cycle*, edited by C. Milas, P. Rothman and D. van Dijk in Elsevier's Contributions to Economic Analysis series.
- Korobilis, D. (2013a). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics* 28, 204-230.

- Korobilis, D. (2013b). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting* 29, 43-59.
- McCracken, M. and Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. Federal Reserve Bank of St. Louis, working paper 2015-012A.
- Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys* 29, 46-75.
- Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* 103, 681-686.
- Raftery, A., Karny, M. and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics* 52, 52-66.
- Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* 72, 821-852.
- Sims, C. and Zha, T. (2006). Were there regime switches in macroeconomic policy? *American Economic Review* 96, 54-81.
- Stock, J. and Watson, M., (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14, 11-30.
- Stock, J. and Watson, M. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking* 39, 3-33.
- Varian, H. (2014). Big Data: New tricks for econometrics. *Journal of Economic Perspectives* 28, 3-28.