Bayesian Compressed Vector Autoregressions

Gary Koop
University of Strathclyde

Dimitris Korobilis
University of Glasgow

Davide Pettenuzzo
Brandeis University

May 28, 2015

Regression models with huge numbers of explanatory variables are now commonly used in various fields (e.g. neuroimaging, molecular epidemiology, astronomy). In economics, similar problems occur with large VARs. Several statistical methods have been developed to deal with the problems which occur when the number of explanatory variables in a regression is much larger than the number of observations. These methods include those which reduce the parameter space through variable selection or coefficient shrinkage (e.g. the LASSO) and those which reduce the dimensionality of the data (e.g. methods such as principal components analysis, PCA, which extract a small number of factors from the huge number of explanatory variables).

In this paper, we wish to investigate an alternative which is growing in popularity: compressed regression through random projections (see Guhaniyogi and Dunson, 2014). We extend the compressed regression methods of Guhaniyogi and Dunson (2014), developed for the regression model with a scalar dependent variable, to the VAR. We begin by outlining the basic ideas and motivation in the regression model before moving on to the VAR.

Let y_t for $t = 1, .., T$ be a scalar observation on a dependent variable which depends on a vector of k explanatory variables, X_t, where $k >> T$ through a regression model:

$$y_t = X_t' \alpha + \varepsilon_t. \tag{1}$$

Working directly with (1) is impossible with some statistical methods (e.g. maximum likelihood estimation) and computationally demanding with others (e.g. Bayesian approaches which require the use of MCMC methods). Some of the computational burden can arise simply due to the need to store in memory huge data matrices. Manipulating such data matrices even a single time can be very demanding. For instance, calculation of the Bayesian posterior mean under a natural conjugate requires, among other manipulations, inversion of a $k \times k$ matrix involving the data. This can be difficult if k is huge. For more complicated models for which MCMC methods are required, similar manipulations must be done for each of $r = 1, .., R$ MCMC replications. Bayesian Compressed Regression (BCR) avoids such demands by not work directly with (1), but rather with the much more parsimonious compressed regression specification:

$$y_t = (\Phi X_t)' \beta + \varepsilon_t \tag{2}$$

where Φ is $m \times k$ and $m << k$. Thus, the k explanatory variables are squeezed into the lower dimensional ΦX_t. For a given Φ, the problem is reduced to the very simple one of estimating a regression with a small number of explanatory variables given by $\widetilde{X} = \Phi X_t$. In this paper Φ is a random projection matrix. There are several methods for generating a random matrix e.g. Φ can be generated from a Normal or Uniform density, or from a sparse random projection scheme that allows for several zeros to occur (see Achlioptas, 2003, for details). The Johnson-Lindenstrauss (1984) lemma shows that a random projection matrix is capable to reduce the data into a lower-dimensional model, while preserving the reconstructive or discriminative properties of the original data. Note that a random projection is data independent and simple to generate. Alternative methods are data-based, for example PCA finds a subspace that maximizes the variance in the data.

The main idea in this paper follows closely ideas in Guhaniyogi and Dunson (2014). That is we compute quickly several such random projections and then update the coefficients of the compressed VAR in equation (2) using conjugate priors which allow for analytical expressions for posteriors and marginal likelihoods (Kadiyala and Karlsson, 1997). Then we average predictions from all different compressed models by their respective marginal likelihoods (i.e. we perform Bayesian Model Averaging).

Such large VARs are empirically interesting to macroeconomists. For example, Banbura, Giannone and Reichlin (2010) estimate a large 132-variable VAR using a Minnesota prior, and they show that it provides superior forecasts compared to typical medium or small VARs; see also Carriero, Clark and Marcellino (2011). In previous work with Gary (Koop and Korobilis, 2013) we have generalized large VARs in order to deal with structural instabilities in its parameters.

While this is work in progress, initial Monte Carlo simulations and runs using real data show very promising results. In particular, the Bayesian Compressed VAR (BCVAR) approach provides in many cases superior forecasts compared to the Minnesota prior and Dynamic Factor Models estimated using principal components.

References

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

[2] Banbura, M., Giannone, D. and Reichlin, L. (2010). “Large Bayesian vector auto regressions,” Journal of Applied Econometrics, 25, 71-92.

[3] Carriero, A., Clark, T. and Marcellino, M. (2011). “Bayesian VARs: Specification choices and forecast accuracy,” Federal Reserve Bank of Cleveland, working paper 11-12.

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

[5] Guhaniyogi, R. and Dunson, D. (2014). Bayesian compressed regression. Journal of the American Statistical Association, DOI: 10.1080/01621459.2014.969425.

[6] Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mapping into Hilbert space. Contemporary Mathematics, 26, 189-206.

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

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