Empirical Project Instructions

Below I give instructions for getting a data set and model and you are asked to: i) describe methods for Bayesian estimation of the model; ii) write or adapt existing code for doing Bayesian estimation; and iii) present empirical results using your code and data set. Beyond this general description, I am deliberately leaving details of the assessment flexible so that you can push your answer in any direction you want and reveal your ability to motivate and explain clearly a piece of Bayesian empirical work. Your written answer should be a maximum of 2500 words plus up to two pages of tables/graphs presenting your empirical results. You should submit your project as a pdf file, sent as an email attachment to me (gary.koop@strath.ac.uk). You should also hand in the computer code and data (zip them together and send as an email attachment to me). You may choose between one of the following two topics.

Topic 1. *The Threshold Autoregressive Model*

This topic relates to the paper Kim, C., Nelson, C. and Piger, J. (2004) "The less volatile U.S. economy: A Bayesian investigation of timing, breadth, and potential explanations," *Journal of Business and Economic Statistics,* 22, 80-93. You may wish to read this paper for a general motivation and example of how researchers present empirical results. If you like, you can work with the model in this paper. However, it uses a class of models (i.e. Markov-switching models) that were not covered in the lectures. Nevertheless, if you are ambitious, you can work with a Markov switching model (contact me and I will suggest some background readings).

Instead I suggest you work with the threshold autoregressive model which is a popular one in macro-econometrics which can be treated as an extension of a regression model. It can be written as:

$$\begin{array}{l} y_t = \beta_{10} + \beta_{11} y_{t-1} + ... + \beta_{1p} y_{t-p} + \sqrt{h_1^{-1}}\epsilon_t \text{ if } z_{t-d} \leq \tau \\ y_t = \beta_{20} + \beta_{21} y_{t-1} + ... + \beta_{2p} y_{t-p} + \sqrt{h_2^{-1}}\epsilon_t \text{ if } z_{t-d} > \tau \end{array},$$

where y_t is the dependent variable and ϵ_t is i.i.d. $N(0,1)$. Note that this model breaks the data into two different regression models, depending on whether z_{t-d} is above or below a threshold τ. You should treat τ as an unknown parameter. The variable z_{t-d} can be an exogenous, deterministic or lagged dependent variable (and you may wish to investigate more than one choice). For instance, if $z_{t-d} = t$ (i.e. it is simply time), then this model becomes one where the model's parameters change a time τ (i.e. it is a type of structural break model). Specifications with $z_{t-d} = y_{t-d}$ have been used to investigate business cycle effects. For instance, if y_t is GDP growth then such a specification has the property that the model changes depending on whether recent GDP growth has been slow (in a recession) or fast (in an expansion). As another example, suppose y_t is GDP growth and z_{t-d} is the oil price (often treated as an exogenous variable), then the economy can exhibit different dynamics depending on whether the oil price is high or low. d is known as the delay parameter and is usually set to a

particular value ($d = 1$). You may wish to work with the basic model I have just described. Possible extensions to this model that you may (or may not) wish to investigate include: i) treating d as an unknown parameter and estimating it, ii) allowing for only the error variance to change (but the AR coefficients to be the same in the two regimes), iii) to allow for more than two regimes or iv) to consider more than one choice for z_{t-d}. I stress I am not requiring you to do all of these suggested extensions (and, given the word limit, you would not be able to do them all in your project). Rather, I want you to use your own empirical intuition and try and decide (for your chosen variable) which (if any) extensions of the basic model are required.

You are asked to describe Bayesian estimation methods for this model, write the relevant code (or adapt code you find on the internet) and carry out a small empirical exercise using a time series variable of your choice. You should find your own time series data set with this project (e.g. from the FRED data base available at http://research.stlouisfed.org/fred2/). You can choose any variable you wish, but a common choice is GDP growth.

Topic 2. *Shrinkage in VARs.*

We discussed VAR models and the computer sessions provided you with some experience at working with these models in practice. For this topic, you should use the code provided in the computer sessions

In the lectures, I argued that VARs can be over-parameterized and that prior shrinkage could be valuable in overcoming this problem. However, my main empirical illustration involved relatively small VARs with only three dependent variables. This question asks you to investigate the role of prior shrinkage using some of the priors discussed in the course in a VAR with more dependent variables. You can choose your own data set of at least 8 dependent variables or you can use the data set described below. Your answer should begin by briefly describing the models and methods you use, before presenting empirical results. As part of the assessment, I leave it to you to decide which empirical results to present (e.g. coefficient estimates or impulse responses) and how to organize their presentation. But you should focus on the issue of how and whether the different priors induce shrinkage. As for which priors you use, I would suggest considering the ones that are already provided in the code (although if you are ambitious you can adapt my VAR code for another prior such as a LASSO prior) but consider different ways of choosing prior hyperparameters. For instance, for the natural conjugate and independent Normal-Wishart priors the code allows you to subjectively choose prior hyperparameter values. You can experiment with different choices for these (e.g. investigate what happens as you move from non-informative to more informative priors) but you could also modify the code to do a training sample prior. The code also uses an SSVS prior and you can consider various choices for its prior.

If you do not have a particular data set of interest to yourself, then you can use data from one of my recent papers which is available at:

http://qed.econ.queensu.ca/jae/2013-v28.2/koop/

If you download the material there you will find a spreadsheet with many US macroeconomic variables. The Data Appendix available there lists the variables, you can use the three variables listed in Table B1 along with five (or more) additional series.

I note that when working with large VARs it is common to transform all variables to stationarity before including them in the VAR. The Data Appendix (in the column labelled Tcode) describes my recommended transformations for each variable.