SGPE ECNM11060

Bayesian Econometrics

Empirical Project

Spring 2026

SCOTTISH GRADUATE PROGRAMME IN ECONOMICS

Semester 2 Options

Dr **Niko Hauzenberger** (point of contact)
niko.hauzenberger@strath.ac.uk

Dr **Ping Wu**
ping.wu@strath.ac.uk

The empirical project must be submitted by **12 noon on 29 May, 2026** electronically to the associated dropbox on Learn. This assessment contributes to **60% of the final mark** for this class.

Names must not appear on assignments, you should only identify yourself on your assignment by your examination number which can be found on your student card. Deadlines are absolute and must be strictly adhered to otherwise the standardised penalty will be applied without exception. Further information on late penalties is available within your Programme Handbook. The penalty will not be applied if good reasons can be given, such as documented illness. If you feel you require an extension, you should contact the Postgraduate Office as soon as possible. The Special Circumstances procedure may be appropriate where a circumstance beyond a student's control has negatively affected their ability to perform or complete a University assessment. For further guidance on the Special Circumstances procedure, please contact the Postgraduate Office, or see your Programme Handbook. When submitting on Learn, your work will be checked by TurnItIn (plagiarism checker). Further guidance on plagiarism, and how to avoid it is available here. In addition, the university's guidelines on the use of generative AI can be found here.

Nature of Assessment and Instructions

Below we give instructions for getting a data set and model and you are asked to:

1. describe methods for Bayesian estimation of the model,
2. write or adapt existing code for doing Bayesian estimation, and
3. present empirical results using your code and data set.

Beyond this general description, we are 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 2,500 words** plus up to two pages of tables/graphs presenting your empirical results. You may choose between one of the following three topics.
The first topic is a new model you will be unfamiliar with (although it is closely related to models covered in the lectures) and we did not provide Matlab code for it in the computer tutorials. You will have to write your own code (or find some on the internet, e.g., at Joshua Chan's webpage). Hence, we would expect more discussion of the econometrics (i.e., describing the form of the model and how the MCMC algorithm works) and your empirical results section can be correspondingly shortened. The second topic involves a more sophisticated model, but it is one covered in the lectures and for which code has been provided (you can even use the BEAR toolbox discussed in the course for this topic). Hence we would expect less discussion of the econometrics (i.e., only a very brief discussion of the model, priors and Bayesian estimation methods based on lecture material), but a more in-depth empirical analysis. The third topic,

involving Bayesian Additive Regression Trees (BART), is the most advanced but if you are interested in microeconomic topics involving cross-sectional data sets is the most relevant. And we offer advice which should greatly simplify the coding process provided you either know R or are willing to spend a small amount of time learning the basics of R.

Topic 1: ***The Moving Average Stochastic Volatility Model.*** This topic relates to methods developed in "Moving Average Stochastic Volatility Models with Application to Inflation Forecasting" by Joshua Chan, published in the *Journal of Econometrics,* 2013, 176(2), 162–172. The definition of the Moving Average Stochastic Volatility (MA-SV) model is given in Eqs. (1) through (3) of this paper. This model is similar to the Unobserved Components Stochastic Volatlity (UC-SV) model that formed the basis of Exercise 1 in Computer Tutorial 4. Both of these are state space models. They differ in that the MA-SV model has a moving average error in the measurement equation whereas the UC-SV model does not. The UC-SV model has stochastic volatility in both measurement and state equations, whereas the MA-SV model only has stochastic volatility in the state equation.

You are asked to describe Bayesian estimation methods for the MA-SV 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 here). You can choose any variable you wish, but a good choice would be a time series of inflation.

We will leave you wide latitude in how you design your empirical exercise, but here we provide some suggestions of issues you might want to explore. As noted in Chan (2013), there are several interesting models nested within the MA-SV model. Thus, model selection issues are relevant. You could estimate a range of models (see Table 1 of the paper) and compare estimates and discuss which model to select. Model selection can be done informally, by looking at parameter estimates in each model (e.g., if the posterior means of MA coefficients are all small relative to their posterior standard deviations then this is evidence against the MA model). More formally, Bayes factors can be calculated comparing models as described in Chan (2013). We note that calculating Bayes factors for these models is more difficult than simply estimating the models so it is not a requirement of this project to do this (but you may wish to do so if you are ambitious).

An alternative issue that could be discussed in this project would be trend inflation. The MA-SV and UC-SV models can both provide estimates of trend inflation. An investigation of how these estimates differ could be an interesting issue to investigate.

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, the BEAR toolbox for Matlab, or the R codes provided here.

In the lectures, we argued that VARs can be over-parameterized and that prior shrinkage could be valuable in overcoming this problem. However, our 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 eight dependent variables or you can use the data set described below (for a reference, see Hauzenberger, Huber, and Koop, 2024). Your answer should begin by briefly describing the models and methods you use, before presenting empirical results. As part of the assessment, we 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, we would suggest considering the ones that are already provided in the code or in the BEAR toolbox (although if you are ambitious you can adapt the VAR code from the lecture 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 can choose some variables from the FRED database. Note that when working with large VARs it is often easier to transform all variables to stationarity before including them in the VAR. The FRED spreadsheets have a row labelled `transform` which describes the recommended transformation for each variable.

Topic 3: *Bayesian Additive Regression Trees (BART).* The recommended reading for BART was Sections 1 and 2 of Hill, J., Linero, A. and Murray, J. (2019). "Bayesian Additive Regression Trees: A Review and a Look Forward" (available on the course website) which covered the basic BART model. The basic BART model was also covered in the lectures and Matlab code was provided in Computer Tutorial 3. Sections 3 through 6 of the Hill *et al.* (2020) paper discuss extensions of the basic BART model. This topic involves either working with the basic BART model or choosing one of these extensions, learning the relevant Bayesian econometric methods and doing some empirical work. Section 7 of the Hill *et al.* (2020) paper describes R software which can be used for the basic BART model as well as its extensions. If you are familiar with R (or are willing to spend a small amount of time to learn the very small amount of R necessary to run the BART software packages), then the coding requirements of this project are small. Sample data sets are typically provided with these software packages and you can use one of them (or you can choose another data set).

There are many BART project topics possible. If you wish to stick with basic BART, then you could do an exercise similar to the "bake-off" in Section 5.1 of Chipman, George and McCulloch (2010). "BART: Bayesian Additive Regression Trees" published in the *Annals of Applied Statistics*. Parts of this paper may be difficult for a student in a first course in Bayesian econometrics (e.g., involving cross-validation which is something we did not cover in the course), but many parts are not and it should not be hard to get a general idea of what their "bake-off" involves. Basically, they take a range of data sets and a range of econometric modelling approaches and sees which works best (in terms of RMSE). You could find a data set and compare BART to other approaches (the simplest choice would be simply to compare BART to a linear regression).

Of the various BART extensions considered in the Hill *et al.* (2020) paper, the BART probit model would be the simplest and you could do a "bake-off" involving this model. If you are interested in binary choice models, this would be an interesting project.

Another option would be to consider causal inference using BART (Section 6 of Hill *et al.*, 2020). Economists are often interested in causality in general and there is a large literature on topics like treatment effects (e.g., the effect of a certain treatment, policy or training programme on the individual receiving the treatment, the latest Nobel Prize winners were experts in this which reveals the interest economists attach to this topic). In the course, we have largely focussed on macroeconomic models and these methods are typically used with microeconomic data (see, e.g., Hahn *et al.*, 2020). Hence, we have said little about causality in the course. Relatedly, Deshpande *et al.* (2024) provide methods (and an **R** package) that allows for the estimation of heterogeneous coefficients, which could potentially be used to model varying treatment effects. Choosing this topic is your chance to learn more about Bayesian methods for investigating causality.

References

Chan JC (2013), "Moving average stochastic volatility models with application to inflation forecast," *Journal of econometrics* **176**(2), 162–172.

Chipman HA, George EI, and McCulloch RE (2010), "BART: Bayesian additive regression trees," *The Annals of Applied Statistics* **4**(1), 266 – 298.

Deshpande SK, Bai R, Balocchi C, Starling JE, and Weiss J (2024), "VCBART: Bayesian Trees for Varying Coefficients," *Bayesian Analysis* **1**(1), 1–28.

Hahn PR, Murray JS, and Carvalho CM (2020), "Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion)," *Bayesian Analysis* **15**(3), 965–1056.

Hauzenberger N, Huber F, and Koop G (2024), "Macroeconomic forecasting using BVARs," in "Handbook of Research Methods and Applications in Macroeconomic Forecasting," 15–42, Edward Elgar Publishing.

Hill J, Linero A, and Murray J (2020), "Bayesian additive regression trees: A review and look forward," *Annual Review of Statistics and Its Application* **7**, 251–278.