Brown University, Department of Economics, Spring 2024

Project for ECON 1670

On project from syllabus (dates are updated)

- A project should consist of a Bayesian analysis of a dataset. It should contain a data description, a model description, model checks, evidence of posterior simulator convergence and implementation correctness, prior sensitivity analysis, and estimation results.
- A project proposal is due on April 25. It should include a data description, a model description, and a description of the MCMC algorithm. The project is due on May 9.

More details

- model checks:
  - prior and posterior predictive analysis, see slides 14-15 in Bayesian modelling and the code of the S&P 500 example on canvas
  - comparison of average predictive logscores for a baseline and more complicated versions of the model; to compute average predictive log score perform the following steps: for each $k \in \{1, \ldots, K\}$, $K = 50$ or 100, randomly select 85% of the sample, estimate the model on this 85% of observations, evaluate predictive log score LS_k on the 15% not used in estimation (formula on p. 30 of Mixtures slides), save the predictive log score LS_k; then, the average predictive logscore is $\frac{1}{K}\sum_{k=1}^{K} LS_k$; if you use time series data, then you could compute a predictive logscore using first 85% observations for estimation and the rest for prediction (so no averaging).
- evidence of posterior simulator convergence - plot marginal posteriors from the first and fourth quarters of MCMC draws, mean equality tests, see the posted code for the normal linear regression example
- evidence of implementation correctness (if you implement an MCMC algorithm yourself): Geweke's joint distributions tests; also, checking that for simulated data the true parameter values are well within the support of the correspondning posteriors distributions, see the posted code for the normal linear regression example
- estimation results: marginal posteriors for parameters of interest; if suitable for your project, provide a solution to a decision problem such as the optimal class size in the tests scores/class size example.
- The use of STAN for the project is encouraged as coding up an MCMC algorithm correctly is time consuming.

Datasets

- It is acceptable to use the datasets posted on canvas: test scores and class size; S&P 500 daily returns.

- Journal of Applied Econometrics has data and replication files for most of the articles: `http://qed.econ.queensu.ca/jae/legacy.html`

- Econometrics textbooks have collections of datasets, e.g.,
  `https://cran.r-project.org/package=wooldridge`
  `https://wps.pearsoned.com/aw_stock_ie_3/178/45691/11696965.cw/index.html`
  `https://gattonweb.uky.edu/sheather/book/data_sets.php`

Specific examples of projects

- A finite mixture of normal distributions for modelling the error term in a linear regression, applied to test scores and class size data, see also Exercise 6.4.4 in Geweke's textbook.
- Application of a Markov mixture model to S&P 500 daily returns data. See, Example 7.3.2 and Exercises 7.3.1-2.