AAEC Bayesian Econometric Analysis

Instructor: Klaus Moeltner

Problem Set 2

General Instructions:
Please type everything in LaTeX (including all Math) and hand in a pdf file. For problems involving Matlab, answer questions in LaTeX, and attach your script, log file, and any graphs to your main pdf file.

If you prefer, you can also load your Matlab figures and/ or output tables into your LaTeX file (see PS1 instructions).

Q1: PPP values for kurtosis
Helpful scripts and functions: `mod3_ppp`

Examine the suitability of our normal linear regression model for the earnings data (script `mod2_application`) based on the kurtosis statistics. The kurtosis is given as

$$k = \frac{N\sum_{i=1}^{n}\varepsilon_i^4}{\left[\sum_{i=1}^{n}\varepsilon_i^2\right]^2} \qquad \text{where} \quad \varepsilon_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}$$

For the normal density, $k = 3$.

Following script `ppp` simulate draws of kurt_y and kurt_ystar, generate a **plot** of their difference, a **plot** of kurt_ystar with E[kurt_y] super-imposed, and compute the PPP value. What is your conclusion regarding the fit of our model to the underlying data based on the kurtosis statistics? (*Hint: You may need to change some of the figure's axes settings to see the full distribution.*)

Label your script, log file and output `ps2_q1`.

Q2: Examining single-valued linear restrictions via HPDIs

Using script `mod4_outage_hpdi` as a template test the following restrictions for the outage data using HPDIs:

1. The cost of a scenario 6 outage for a size 3 firm is no larger than the cost of the same outage to a size 1 and a size 2 firm combined.
2. The cost of a scenario 5 outage to a size 2 firm is exactly twice the cost of the same outage to a size 1 firm.
3. The differences in cost between outage scenarios 5 and 6 across all firm sizes sum to zero.

For each test create a posterior plot with super-imposed HPDI bounds and the actual numerical values of these bounds, just like in the class script. (*Hint: You may need to change some of the figures axes settings to see the full distribution*.) Label your script, log-file, and output as `ps2_q2`. **State your conclusion for each "test."**

Q3: Examining multi-valued linear restrictions via SDDR

Using script `mod4_outage_sddr` as a template examine the following restrictions for the outage data:

For scenario 1, the cost for size 3 is the sum of cost for size 1 and cost for size 2 AND for scenario 6, the cost for size 3 is twice the sum of the costs of the other two sizes.

a) Express your restriction as $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$ and show the explicit form of **R** and **q** in your script.
b) Compute the log-SDDR value. What is your conclusion regarding the likelihood of this hypothesis?

Label your script, log-file, and output as `ps2_q3`.
(Note: Your script will take a few minutes to execute)

Q4: Examining restrictions via the Chib Method

Consider our wage regression for the normal regression model with independent priors (Matlab script `mod2_application`, function `gs_normal_independent`).

a) Re-run the script `mod2_application.` Label your script, log, and output as "`ps2_q4_M1`". Derive the log-marginal likelihood value for this model via Chib's method by following these steps:

- Create a starter script "`ps2_q4_M1_chib`", using `mod4_sur_chib` as an example. Label your log file and output "`ps2_q4_M1_chib`" as well. You need to alter `mod4_sur_chib` in several ways:
  - Load in the correct original data and the correct priors from the main script;
  - Load in the output from `ps2_q4_M1`
  - Call function "`gs_normal_independent_chib`" with the correct input arguments (the output arguments will remain unchanged).
- Create function `gs_normal_independent_chib.` Use function `gs_sur_chib` as a template and make all necessary adjustements. You will need to evaluate the inverse gamma density. The function for this is posted on the web (`invgampdf`).
- Run script `ps2_q4_M1_chib`. If all goes well, your log file should show your log-prior, log-likelihood, log-posterior, and log-marginal likelihood evaluated at your posterior parameter means.

b) Save `ps2_q4_M1` as `ps2_q4_M2` and alter it as follows: Add the following regressor to your original model: husband's hours worked in 1975 **in log form**, husband's age, and husband's educational attainment. Label your log file & output as `ps2_q4_M2` and run the script.

   Examine your posterior output. How would you describe the posterior density for the added coefficients? Do these variables seem to have much explanatory power?

c) Create a starter script for this model ("Model 2") by adjusting `ps2_q4_M1_chib` accordingly. Label this script, its log file and its output `ps2_q4_M2_chib.` Run the script.

   Compare the sets of results for `ps2_q4_M1_chib` and `ps2_q4_M2_chib`.

   i. What is the log-Bayes Factor for M1 over M2?
   ii. Which model thus receives stronger support by our data?
   iii. Looking at the detailed results, which component entering the computation of log-mLH appears to be the primary driver of the difference in mLH's between the two models?

iv. What does this suggest regarding the effect on the mLH of added parameters with diffuse priors?

v. In this context, does the mLH "reward" sparseness in parameters?

d) Now examine the linear restriction implied by M1 as compared to M2 using the SDDR approach.

a. Express your restriction as $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$ and show the explicit form of **R** and **q** in your script.
b. Compute the log-SDDR value. Does your result (approximately) agree with result from Chib's method? Label this script and its log file "`ps2_q4_sddr`".