Bayesian Econometrics: Meet-up 2

Kevin Arthur Tittel

February 17, 2023

1 Recap Meet-up 1

In this Section we briefly review some main take-aways from Meet-up 1. We introduced the Bayesian framework, with a range of advantages over the frequentist approach in estimating econometric models (e.g. exact results in small samples, more suitable than ML for hierarchical statistical models, natural way to deal with uncertainty and incorporating prior knowledge, optimal for decision making under uncertainty subject to certain conditions). Instead of assuming a fixed parameter vector θ and one realization of (a possibly small sample of) data y, Bayesian analysis changes perspectives (from a subjective point of view) by allowing uncertainty in θ through a probability distribution (*True or False Question 1*). The more informative the prior distribution, the greater its influence on the posterior distribution (*True or False Question 2*), such that the likelihood function overtakes the posterior distribution in case of uninformative priors. Seen differently, very informative priors allow for accurate results in case only a small sample y is available. In a similar way of reasoning, we need a larger sample size in case the prior is weak (*True or False Question 3*).

The posterior distribution of parameters θ can be observed from the kernel function without specific knowledge on the value of (redundant) constants, whereas the full specification (with constant c is required in case we want the posterior *function* value for a particular value of θ (*True or False Question 4*). In case of a linear regression models, we are interested in the posterior distribution of $\theta = [\beta, \sigma^2]|y$. That is, we start off with solely prior knowledge on $\theta = [\beta, \sigma^2]$ (i.e. often $p(\beta, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2)$), after which we start adding data y through the likelihood function $p(y|\beta,\sigma^2)$, eventually ending up with the posterior distribution through $p(\beta,\sigma^2|y) \propto p(\beta|\sigma^2)p(\sigma^2)p(y|\beta,\sigma^2)$. In linear regression models, the posterior distribution is of the same family as the prior distribution (i.e. normally distributed) in case we choose $\beta|\sigma^2 \sim N(b,\sigma^2 B)$ and $p(\sigma^2) \propto \sigma^{-2}$ (*Extra Exercise 18*) or $\sigma^2 \sim IG-2(\delta,\nu)$ (*Extra Exercise 13*). In other words, the chosen priors are conjugate priors (*True or False Question 5*). The more precise these priors are (i.e. the smaller their variance), the larger of an influence they have on the posterior distributions (e.g. posterior mean) (*True or False Question 6*). In general, with an increasing number of observations the priors' influence becomes less since the likelihood takes up a larger part in the posterior, but there are certain exceptions (*True or False Question 7*, *Extra Exercise 18b*).

2 Exercise

Consider the linear regression model $y = X\beta + \epsilon$ for a data set on standardized SAT scores for $N = 10,000$ potential university students, where $y = [y_1, ..., y_N]'$, $X = [x_1, ..., x_N]'$, x_i being a vector of explanatory variables of dimension $k = 10$, and $\epsilon = [\epsilon_1, ..., \epsilon_N]' \sim N(0,\sigma^2 I_N)$. We perform a Bayesian analysis with $\beta|\sigma^2 \sim N(0,\sigma^2 B)$ and $\sigma^2 \sim IG-2(0,\nu)$ [use that $p(\sigma^2) \propto (\sigma^2)^{-\frac{\nu+2}{2}}\exp(-\frac{\delta}{2\sigma^2})$].

a. Derive the joint posterior distribution, i.e. provide the kernel/density of $p(\beta,\sigma^2|y)$. What is the name of this distribution and what are the parameters?

b. Derive the joint posterior distribution, i.e. provide the kernel/density of $p(\beta|y)$. What is the name of this distribution and what are the parameters?

c. Derive the joint posterior distribution, i.e. provide the kernel/density of $p(\sigma^2|y)$. What is the name of this distribution and what are the parameters?

3 True or False Questions

1. In case we want to obtain the marginal posterior density $p(\beta|y)$, we first need to establish the joint posterior density function $p(\beta,\sigma^2|y)$, after which we integrate this function over σ^2.

2. The joint posterior density function $p(\beta,\sigma^2|y)$ is always proportional to the function $p(\beta)p(\sigma^2)p(y|\beta,\sigma^2)$.

3. In a Bayesian analysis framework, researchers can iteratively include (updated) "prior beliefs" on parameter estimates in an iterative estimation procedure.

4. In linear regression models, one may choose to only impose uncertainty on β, whilst specifying a (known) value for σ^2 (i.e. $p(\beta, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2) = p(\beta)$ in case $\sigma^2 = c \in \mathbb{R}$).

5. *Lecture 2 pg. 11-13.* The tables and graphs of posterior results confirms that, given a fixed sample size, a more precise prior i) affects the posterior mean results to a lesser extent and ii) decreases the posterior scale parameter estimates.

6. In a (simple) binary choice model where the random variable Y follows a binomial distribution (i.e. $P[Y_i = 1] = \lambda$ for $i = 1, ..., N$ with $0 < \lambda < 1$) and the posterior distribution follows a Beta distribution, a conjugate prior on λ i) is also a Beta distribution and ii) decreases the degrees of freedom of the posterior distribution compared to when one were to choose for a diffuse prior (i.e. $\lambda \sim U([0, 1])$).

7. *Lecture 2 pg. 19-21.* The tables and graphs of posterior results confirms that, given a fixed sample size, a prior with larger variance affects the posterior density to a larger extent.

8. When presenting Bayesian results on the posterior distribution of θ, i) certain point estimate results for θ can be computed depending on the desired loss function, and ii) uncertainty can be computed using the posterior variance.

9. In case we test for $\theta = \theta_0$, our parameter θ is estimated to be significantly different from θ_0 in case $\hat{\theta}$ lies outside the $100(1-\alpha)\%$ highest posterior density region around θ_0.

10. For formal Bayesian testing, so-called Bayes factors can be used (i.e. posterior odds, or "Bayesian likelihood ratio tests"), for which we require i) models to be nested, and ii) the marginal likelihood $p(y) = \int p(y|\theta)d\theta$.

4 Detailed Derivation Questions

- *Extra Exercise 13a.* Why does the OLS standard error $\tilde{\sigma}^2$ of the regression of $w = [y, B^{-1/2}b]$ on $V = [X, B^{-1/2}]$ contain i) the division by $N + \nu$ and ii) the additive term of δ?

- *Extra Exercise 13b.* Why does the integral disappear in the penultimate step of the derivation? Give the distribution of β.