Bayesian Econometrics: Meet-up 3

Kevin Arthur Tittel

March 2, 2023

1 Recap Meet-up 2

In this Section we briefly review some main take-aways from Meet-up 2. We elaborated on the concepts, notation, and mathematics introduced in Meet-up 1, in particular applying our knowledge in an exercise including a linear regression model with conjugate priors. That is, we derived the joint posterior distribution of $\beta, \sigma^2|y$, and the marginal posterior distributions of $\beta|y$ and $\sigma^2|y$, in a linear regression model with conjugate priors for $\beta|\sigma^2$ (normal) and σ^2 (inverse gamma-2). Since these are informative, conjugate priors for a linear regression model (where the likelihood is normal), the joint posterior density is in the same family as the joint prior, which boils down to a normal distribution. Subsequently, the marginal posterior densities of β (multivariate t) and σ^2 (inverted gamma-2) are obtained by integrating the joint posterior density function $p(\beta, \sigma^2|y)$ over σ^2 and β, respectively (*True or False Question 1*). An exact specification of $p(\beta, \sigma^2|y)$ can be determined using the priors and likelihood, i.e. $p(\beta|\sigma^2)p(\sigma^2)p(y|\beta, \sigma^2)$, where $p(\beta|\sigma^2) = p(\beta)$ if β and σ^2 are independent (*True or False Question 2*). The latter could be the case if a researcher chooses to fix a constant, known value $c \in \mathbb{R}$ for σ^2, hence only imposing uncertainty on β such that $p(\beta, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2) = p(\beta)$ (*True or False Question 4*).

In general, given a fixed sample size, a more precise prior (i.e. with a smaller variance due to more certainty in the prior beliefs of the estimates) affects the posterior mean to a larger extent, and decreases the posterior scale parameter estimates due to less uncertainty (*True or False Question 5, 7*). Next to the linear regression model with conjugate priors, another well-known combination of distributions belongs to the (simple) binary choice model where: the prior is beta and the likelihood is binomial, after which the posterior turns out to be beta as well. In such cases, an informative prior increases the number of degrees of freedom of the posterior distribution, compared to when one were to choose for a diffuse prior (*True or False Question 6*). Given this fundamental knowledge on Bayesian analysis, we can move on to presenting Bayesian results of interest, and perform formal and informal Bayesian tests, as we will discuss in Meet-up 3.

2 Discussion Theory

In this Section we start off with a brief description of the theory given the slides of Lecture 2, which we will subdivide into a selection of key concepts. The majority of Lecture 2 concerns comparing different Bayesian models, which implies comparing models and priors rather than testing parameter restrictions. A good thing to keep in mind is that, as usual, such results heavily rely on prior specification (instead of choosing a frequentist approach-alike significance level). The ways of comparing models in Lecture 2 are:

- Highest Posterior Density [HPD] interval/region tests (informal);
- Bayes factor for (non-)nested models using marginal likelihoods (formal);
- Posterior odds analysis (formal);
  - Compute posterior model probabilities;
  - Compute weighted posterior results (e.g. posterior predictive densities);
- Bayes factor for nested models using Savage-Dickey Density Ratio (formal);
- Predictive accuracy, i.e. compute RMSE, MAE, etc. based on point forecasts (mean by quadratic loss, median by absolute loss, mode by zero-one loss of the forecast distribution);
- Predictive power, i.e. predictive odd ratios (also in case of diffuse priors).

How do we obtain Bayesian point estimates? Instead of reporting entire prior and posterior densities of the model parameters θ, for the matter of feasibility, we will often choose to report point estimates of θ, such as the mean, mode, or median of the posterior distribution or posterior predictive density. Depending on the loss function that a researcher chooses, the best, or optimal, point estimate θ_p given the data y can be obtained by minimizing the expected loss, whilst taking this expectation with respect to the posterior distribution (i.e. $\mathbb{E}_{\theta|y}[L(\theta_p|\theta)|y]$ with $L(\theta_p|\theta)$ the loss function in θ_p given the true parameter values θ). All in all, a quadratic loss function yields the mean, the best point estimate of the absolute loss function is the median, and the optimal point estimate of the zero-one loss function is the mode of the posterior density.

How do we perform Bayesian testing? Since Bayesian analysis does not condition on repeated sampling, frequentist testing principles (e.g. $t-$, $F-$, LM-, or Wald tests) cannot be used. Instead, either formal (Bayes factors or posterior odds) and informal approaches (Highest Posterior Density [HPD] interval/region tests) can be used. Formal results are mostly based on marginal likelihoods of two (non-)nested models, which can be computed analytically or using an integral. In case two models are in fact nested, a short-cut, Savage-Dickey Density Ratio can be computed. Subsequently, these Bayes factor results can be used in posterior odds analysis, in which models also get attached a prior belief probability, and with which posterior probabilities or weighted posterior results can be computed.

How do we make Bayesian forecasts? Bayesian forecasts trivially accounts for the parameter uncertainty in θ. That is, we compute the so-called posterior predictive density of y_{N+1}, denoted by $p(y_{N+1}|y)$, by integrating the function $p(y_{N+1}|\theta)p(\theta|y)$ over θ. For the matter of reference, in a frequentist approach one simply imputes the estimated value $\hat{\theta}$ in $p(y_{N+1}|\theta)$. Subsequently, similar to the Bayes factor and posterior odds analysis, predictive odds could be computed, where the ratio of predictive likelihoods of two models is obtained in out-of-sample observations, that is, $p_A(y_{N+1}, ..., y_{N+L}|y)/p_B(y_{N+1}, ..., y_{N+L}|y)$.

3 True or False Questions

1. 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.

2. In case we test for $\theta = \theta_0$, our parameter θ is estimated to be significantly different from θ_0 in case θ_0 lies outside the $100(1-\alpha)\%$ highest posterior density region of the posterior distribution $\theta|y$.

3. 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. If the Bayes Factor $BF_{A|B} = p_A(y)/p_B(y)$ exceeds 1, model A is *a posteriori* more likely than model B, given that proper priors are used on parameters θ with an equal amount of free parameters or restrictions, implying that *a priori* both models are assumed to be equally likely.

5. The Posterior Odds ratio $PO_{A|B} = \Pr[A]/\Pr[B]$ is defined as the product of the prior odds ratio multiplied with the Bayes Factor, and can be used to construct posterior probabilities $\Pr[A|y]$ and $\Pr[B|y]$ in order to select a model with the highest posterior probability.

6. In a simple binary choice model, we can compute the marginal likelihood by either i) solving the integral $p(y) = \int p(y|\lambda)d\lambda$ (and using that the density function of a beta distribution integrates to 1), or ii) using the relation $p(y) = p(\lambda|y)/p(\lambda)p(y|\lambda)$.

7. When performing Bayesian forecasting, in contrast to the classical approach, one fully takes into account the parameter uncertainty in θ by computing the posterior predictive density of $p(y_{N+1}|y) \propto \int_{-\infty}^{\infty} p(y_{N+1}|\theta)p(\theta|y)d\theta$.

Update: March 15, 2023 In the True or False Question 2, it was originally stated that $\hat{\theta}$ should be outside the $100(1-\alpha)\%$ highest posterior density region around θ_0. However, the reasoning should be that the parameter restriction of interest, $\theta = \theta_0$, should be outside the HPD of the posterior distribution $\theta|y$. Note that the question was intended to be false in the first place, but solely because of the word "significantly" (we don't use that word in Bayesian analysis, only "enough posterior support that..."). I adjusted the question because i) HPDs are not built around the true value θ_0 (but merely an automatic result of the entire posterior density function $p(\theta|y)$, and ii) we don't compute any point estimate $\hat{\theta}$ in HPD testing, but we compare the value of interest $\theta = \theta_0$ to the computed HPD.