Bayesian methods A Theory and its Significance: A View

Afaq Ahmad, Dept. of Statistics, OPJS University, Churu (Rajasthan)
Dr. R. C. Sharma, Professor, Dept. of Statistics, OPJS University, Churu (Rajasthan)

Introduction

Over the course of the past decade, the use of Bayesian approaches in the field of econometrics has become an increasingly significant component. There is little question that this is being fueled by a growing recognition of the benefits that Bayesian inference provides. In particular, it gives us a formal mechanism to include the previous information that we frequently hold before viewing the data, it fits in nicely with sequential learning and decision making, and it immediately leads to exact small sample findings. All of these benefits come as a result of the fact that it was developed. In addition, the Bayesian paradigm is especially well-suited for prediction since it takes into consideration the uncertainty of any and all parameters, as well as the model itself. The predictive distribution is the sampling distribution in which the parameters are integrated out with the posterior distribution. The predictive distribution is precisely what we need for predicting, which is frequently one of the primary goals of time-series research. In most cases, the selection of a specific econometric model is not prespecified by theory, and the consideration of a wide variety of competing models is possible. The so-called posterior odds, which are calculated by multiplying the prior chances by the Bayes factor, are one way that a formal comparison of models may be carried out inside a Bayesian framework. The Bayes factor between any two models is the ratio of the likelihoods integrated out with the appropriate prior and summarises how the data favour one model over another. Bayes factors are calculated using the likelihoods integrated out with the prior. This automatically leads to posterior model probabilities whenever there is a collection of plausible models to consider. A natural technique to deal with model uncertainty is to utilise the posterior model probabilities to average out the inference (on observables or parameters) pertaining to each of the different models rather than selecting a single model. This is preferable to the alternative of selecting a single model. The term for this technique is the "Bayesian model average." The latter was first stated in Leamer (1978), and recent applications to economic issues include Fernandez' et al. (2001) for growth regressions, Garratt et al. (2003) for macroeconomic forecasting, and Jacobson and Karlsson (2004) for microeconomic forecasting. The specification of prior distributions for all parameters in the model that are viewed as unknown is an obligatory prerequisite for utilising the Bayesian paradigm. This is an unavoidable need for adopting the Bayesian paradigm. This topic has been the subject of significant dispute, with the argument about whether prior should be applied to the coefficients of basic autoregressive models serving as an excellent illustration of this. When seen from a sampling theoretical vantage point, the problem of testing to see if there is a unit root (determining whether or not to difference the series before modelling it through a stationary model) is fraught with a great deal of difficulty. Comparing models in terms of their posterior chances offers a highly natural Bayesian approach to testing that does not rely on asymptotics or approximations. This method can be found in the paper "Comparing Models in Terms of Their Posterior Odds." It is, of course, sensitive to how the competing models are defined (for example, should we contrast the stationary model with a model that has a root that is greater than or equal to one?), as well as to the choice of prior. The latter concerns have resulted in significant debate within the academic community, and as a result, a special issue of The Journal of Applied Econometrics was published with a lively discussion centering on the chapter written by Phillips (1991). The last chapter argued in favour of employing Jeffreys' principles as a way to reflect previous ignorance of the parameters (for further discussion on this topic, see Chapter 6 of Bauwens et al., 1999). Prediction, much like the selection of one model over another competing model, may also be significantly impacted by the prior. In point of fact, prediction is frequently much more sensitive than parameter inference to the choice of priors (especially on autoregressive coefficients). Furthermore, Koop et al. (1995) demonstrate that imposing stationarity through the prior on the autoregressive coefficient in a simple AR(1) model does not necessarily lead to stabilisation of the predictive variance as the forecast horizon increases.

Bayesian Computational Method

The availability of very efficient and flexible algorithms for conducting inference through simulation in combination with ever more powerful computing facilities has made the Bayesian analysis of non-standard problems an almost routine activity. This has contributed to the increased use of Bayesian methods in econometrics, which is at least in part a consequence of the availability of these algorithms. In particular, the application of Markov chain Monte Carlo (MCMC) methods has resulted in the creation of a highly helpful class of computer algorithms and has sparked a true revolution in the process of using Bayesian approaches. Bayesian inference, which before to the year 1990 was at best a challenging task in practise, reserved for a few number of specialised researchers and limited to a pretty restricted range of models, has now become a highly accessible method that can reasonably simply be used to nearly any model. This is in contrast to the situation before 1990, when it was limited to a very restricted set of models and was reserved for a small number of specialised researchers. The notion behind Markov chain Monte Carlo (MCMC) methods is that inference may be made about an analytically intractable posterior (typically in high dimensions) by producing a Markov chain that converges to a series of drawings from the posterior distribution. This is the primary concept behind MCMC methods. Once one has such a chain of drawings, one can obviously instantly engage in predictive inference as well due to the availability of this information. It is possible to create a Markov chain in a number of different ways, the specifics of which are determined by the nature of the problem. The Gibbs sampler and the Metropolis Hastings sampler are the ones that are utilised the most frequently. It is possible to simplify the implementation of the MCMC sampler by making use of data augmentation, which involves the addition of auxiliary variables to the sampler. As a result, the analysis is often carried out on an augmented space that incorporates not only the model parameters but also items like latent variables and missing observations. Gamerman (1997) is an example of a resource that is easily accessible and discusses MCMC approaches. As a result of this, we are now able to undertake Bayesian analysis of time series models that have been around for a long time (such as ARMA models), as well as of more recent additions to our library of models, such as Markov switching and nonparametric models, and the literature is quite extensive. This is made possible by the fact that Bayesian analysis can now be applied to models that have been around for a long time. As a result, I will have to exercise discretion and will make an effort to focus on a chosen number of topics that I consider to be of great importance. I am hoping that this can provide some insight into the potential function that Bayesian approaches can fulfil in contemporary time series analysis.

Scope of study

This model has the potential to be extremely supporting and may be utilised to obtain the requisite performance of Bayesian approaches in the statistical analysis of time series. Even the most complicated time series models are now possible to be analysed using Bayesian methodology thanks to the application of Markov chain Monte Carlo techniques. Models such as ARIMA models and their fractionally integrated equivalents, state-space models, Markov switching and mixing models, and models that allow for time-varying volatility are some of the models that are explored in some length in this chapter. A summary of various current techniques to nonparametric Bayesian modelling of time series is presented in the concluding section.

Examination

Data called local field potentials, or LFPs, were gathered from rats in order to investigate the neural activity that is involved in feature binding. Feature binding is the process by which the brain combines several types of sensory information into a single neural representation (Botly and De).

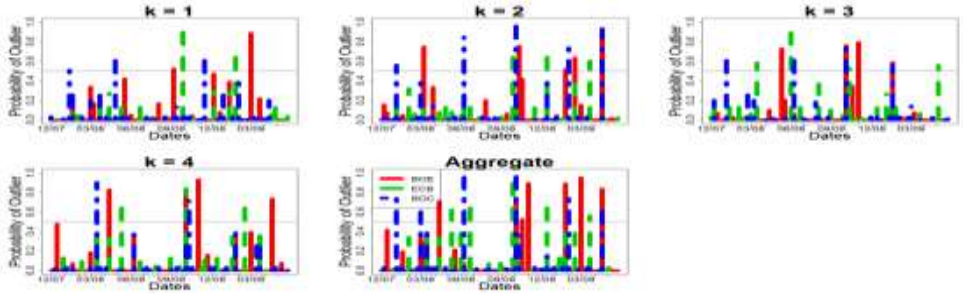
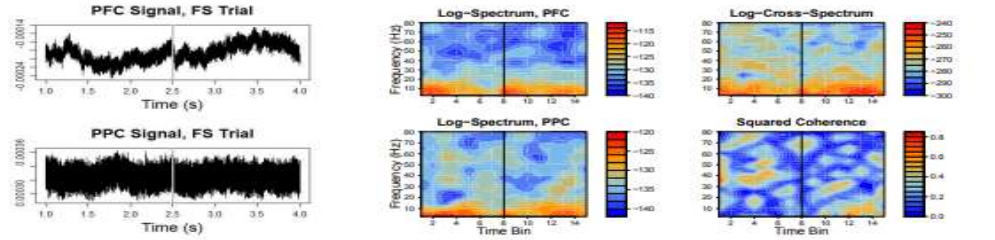


Figure: The MCMC sample proportions of $r^2_{k,(c),t}$ and $\sum_{k=1}^{4} r^2_{k,(c),t}$ hat exceed the 95th percentile of the assumed χ 2 -distributions.

Rosa, 2009; Ljubojevic et al., 2013). For the purpose of recording neural activity throughout time, LFP makes use of pairs of electrodes that are implanted directly in local brain areas of interest. In this instance, the prefrontal cortex (PFC) and the posterior parietal cortex (PPC) are the brain regions that are of interest. The rats were given two different sets of tasks to complete: one set needed them to process numerous stimuli in order to get a reward (known as feature conjunction, or FC), while the other set only required them to process a single stimulus in order to obtain a reward (known as feature singleton, or FS). Both sets of activities were designed to test the rats' ability to recognise patterns in the stimuli they were exposed to. In FC, feature binding is performed, but in FS, a baseline may be established. During each of the tasks, which consisted of 20 separate trials for the FS and the FC, electrodes that had been implanted in the PFC and the PPC monitored the activity of the neurons. As a result, the raw data signal is a bivariate time series with 40 replications for each rat. In Figure a, we present an illustration of the bivariate signals for one of the replications. We will refer to the behavior-based laboratory estimate of the time at which the rat processed the stimuli as t. Each signal replicate lasts for three seconds, and the length of the signal has been centred around this estimate.



(a) The bivariate LFP signal. (b) The associated (log-) spectra and squared coherence.

Figure: The unprocessed LFP data collected from a rat during an FS experiment. The estimated amount of time, denoted by t, that the rat spent processing the stimuli is indicated by the vertical lines.

The behaviour of these bivariate signals as it varies over time, as well as the interplay between them, is of particular interest to us. Utilising time-frequency analysis is an obvious way to address the problem; however, accurate inference for typical time-frequency processes is not yet accessible. Using time-frequency approaches to turn the bivariate signal into an MFTS, which makes available the multivariate modelling and inference capabilities of the MFDLM, is an option that has a lot of potential and should be considered. Since the MFDLM offers smoothing in both the frequency domain T and the time domain T, we may employ time-frequency preprocessing that only delivers a little amount of smoothing since it operates in the time domain T. In the time domain, the signal is divided into time bins with a width equal to one-eighth the duration of the original signal, and the bins that are next to one another have an overlap of fifty percent in order to avoid unwanted boundary effects. Within each time bin, we compute the periodograms of the bivariate signal as well as the cross-periodogram of the signal. Let $q_t^{(1)}(\tau)$ and $q_t^{(2)}(\tau)$ be the discrete Fourier transforms of the PFC and PPC signals, respectively, for time bin t evaluated at frequency τ , after removing linear trends. The periodograms are $I_t^{(c)}(\tau) = |q_t^{(c)}|^2$ for $c = 1, 2$ and the cross periodogram is

$I_t^{(3)}(\tau) = q_t^{(1)}\bar{q}_t^{(2)}$, where $\bar{q}_t^{(2)}$ is the complex conjugate of q (2) t . The cross-periodogram is generally complex-valued, and if the periodograms are unsmoothed, then $|I_t^{(3)}(\tau)|^2 = I_t^{(1)}(\tau)I_t^{(2)}(\tau)$ is real-valued, yet it is abundantly evident that it does not offer any novel information (Bloomfield, 2004). This does not indicate that the cross-periodogram is not relevant; rather, it suggests that the periodograms require some frequency domain smoothing. To generate the smoothed periodograms, often known as spectra, we use a modified version of the Daniell kernel, as recommended by Shumway and Stoffer (2000). Each time bin is then divided into five segments, and the totals are computed. $I_t^{(c)}(\tau), c = 1, 2, 3$ within each segment, and then average the resulting periodograms using decreasing weights determined by the modified Daniell kernel. Denoting these spectra by $\tilde{I}_t^{(c)}(\tau)$, we let $Y_t^{(c)}(\tau) = \log\left(\tilde{I}_t^{(c)}(\tau)\right)$ for $c = 1, 2,$ When the log transformation is advantageous since it is the variance-stabilizing transformation for the periodogram (Shumway and Stoffer, 2000). This makes the log transformation an interesting option. One option available to take into consideration the periodic dependency that exists between signals is the log-cross-spectrum, log. $\left(|\tilde{I}_t^{(3)}(\tau)|^2\right)$. An appealing alternative is the squared coherence defined by $\kappa_t^2(\tau) \equiv |\tilde{I}_t^{(3)}(\tau)|^2/(\tilde{I}_t^{(1)}(\tau)\tilde{I}_t^{(2)}(\tau))$, which satisfies the constraints $0 \leq \kappa_t^2(\tau) \leq 1$ and is the frequency domain analog to the squared correlation (Bloomfield, 2004). Since specifies that $Y_t^{(c)}(\tau) \in \mathbb{R}$, we transform the squared coherence and let $Y_t^{(3)}(\tau) = \Phi^{-1}(\kappa_t^2(\tau)) \in \mathbb{R}$, where $\Phi^{-1} : [0, 1] \to \mathbb{R}$ is a known monotone function; we use the Gaussian quantile function. We have found that fitting $Y_t^{(3)}(\tau)$ produces very similar results to fitting κ 2 t (τ) directly, yet in the transformed case, our estimate of the squared coherence $\Phi\left(\mu_t^{(3)}(\tau)\right)$ complies with the requirements. As a result of the Bayesian methodology that we employ, this modification does not prevent inference.

Distribution of Bayesian Method:

In a broader sense, this method may be used to '-dimensional time series, which, when combined with either the squared coherence or the cross-spectra, results in a C = '('+1)/2-dimensional multifractal time series (MFTS). In Figure b, we present an example of the MFTS that was produced by a rat when it was participating in an FS experiment. In the sake of thoroughness, we have also included the logcross-spectrum, even though it is not a part of the MFTS.

We employ the typical FLCs model described in Section in conjunction with a random walk model for the factors, which are as follows:

$$\begin{cases} Y_{i,s,t}^{(c)}(\tau) = \sum_{k=1}^{K} \beta_{k,i,s,t}^{(c)} f_k(\tau) + \epsilon_{i,s,t}^{(c)}(\tau), & \left[\epsilon_{i,s,t}^{(c)}(\tau)\middle|\sigma_{(c)}^2\right] \overset{indep}{\sim} N(0, \sigma_{(c)}^2) \\ \boldsymbol{\beta}_{k,i,s,t} = \boldsymbol{\beta}_{k,i,s,t-1} + \boldsymbol{\omega}_{k,i,s,t}, & \left[\boldsymbol{\omega}_{k,i,s,t}\middle|\mathbf{W}_k\right] \overset{indep}{\sim} N(\mathbf{0}, \mathbf{W}_k) \end{cases} \quad (2.9)$$

where $\boldsymbol{\beta}_{k,i,s,t} = (\beta_{k,i,s,t}^{(1)}, \ldots, \beta_{k,i,s,t}^{(C)})'$, $Y_{i,s,t}^{(c)}$ are the log-spectra for c = 1, 2 and the probit-transformed squared coherences for c = 3, i = 1, . . . , 8 index the rats, s = 1, . . . , 40 index the trials for each rat, and t = 1, . . . , 15 index the time bins for each trial. The joint indices (i, s, t) in correspond to the time index t in and are used to specify independence of the residuals ωk,i,s,t between rats and between trials. For each initial time bin $t = 1$, we let $\boldsymbol{\beta}_{k,i,s,1} \sim N(\mathbf{0}, 10^4\mathbf{I}_{C\times C})$, since the corresponding observations are only time-ordered within a trial. The C × C factor covariance matrices Wk do not depend on the rat or the trial, and can help summarize the overall dependence among factors. For simplicity and

parsimonious modeling, assumes independence between ωk,i,s,t and ωj,i,s,t for $j \neq k \in \{1, \ldots, K\}$, For the error variances, we use the conjugate priors $\sigma_{(c)}^{-2} \overset{iid}{\sim} \text{Gamma}\,(0.001, 0.001)$ and $\mathbf{W}_k^{-1} \overset{iid}{\sim} \text{Wishart}((\rho R)^{-1}, \rho)$, with R−1 = IC×C, the expected prior precision, and ρ = C ≥ rank(R−1). We provide the full conditional posterior distributions in Appendix A.

To determine the effects of feature binding, we compare the values of $\mu_{i,s,t}^{(c)}(\tau)$ during the interval between the FS and FC trials. We estimate posterior distributions for the sample means by letting Si,F C (respectively, Si,F S) be the subset of FC (respectively, FS) trials for which rat i got the reward. This allows us to calculate the posterior probability of each sample mean.

$$\bar{\mu}_t^{(c)}(\tau) \equiv \frac{1}{8}\sum_{i=1}^{8}\left[\frac{1}{|S_{i,FC}|}\sum_{s\in S_{i,FC}}\mu_{i,s,t}^{(c)}(\tau) - \frac{1}{|S_{i,FS}|}\sum_{s'\in S_{i,FS}}\mu_{i,s',t}^{(c)}(\tau)\right] \text{ for } c = 1,2 \text{ and}$$

$$\bar{\mu}_t^{(3)}(\tau) \equiv \frac{1}{8}\sum_{i=1}^{8}\left[\frac{1}{|S_{i,FC}|}\sum_{s\in S_{i,FC}}\Phi\left(\mu_{i,s,t}^{(3)}(\tau)\right) - \frac{1}{|S_{i,FS}|}\sum_{s'\in S_{i,FS}}\Phi\left(\mu_{i,s',t}^{(3)}(\tau)\right)\right].$$

We investigate the difference in the log-spectra and the squared coherences between the FC trials and the FS trials by taking an average of the results across all rats and across all trials in which the rat responded appropriately to the stimuli. This restriction is significant because it eliminates trials that are not typical of the population being studied, in particular FC trials in which feature binding may not have taken place.

Conclusions

A broad framework for modelling complicated dependency among functional data is provided by the MFDLM. Because we isolate the functional component by means of the proper conditioning and incorporate the requisite identifiability requirements, we are able to model the remaining dependency utilising scalar and multivariate modelling techniques that are already well known to us. Using the hierarchical Bayesian method, we are able to incorporate interesting and helpful submodels in a seamless manner. Some examples of these are the common trend model described in Section, the stochastic volatility model described in Section, and the random walk model described in Section We represent the functional component as a collection of smooth and optimum curves that are constrained by (identifiability) requirements using a combination of Bayesian spline theory and convex optimisation. We execute inference on any parameters of interest, such as (c) t in the LFP example, by obtaining posterior samples of all of the unknown parameters in using an efficient Gibbs sampler. This enables us to perform inference on any parameters of interest. The adaptability and versatility of our technology is demonstrated by the two distinct applications that we have developed. The common trend model presented in Section offers helpful insights into the interconnections between the yield curves of many economies, and the LFP example shown here gives a unique method to time-frequency analysis through the application of MFTS. In these applications, the MFDLM accurately models a variety of functional dependence structures, such as time dependence, (time-varying) contemporaneous dependence, and stochastic volatility. Additionally, the MFDLM is able to readily accommodate additional dependence structures, such as covariates, repeated measurements, and spatial dependence. An R package based on our methodologies is currently in the process of being developed.

Referneces

[1] Lamb, H. H. 1972 British Isles weather types and a register of the daily sequence of circulation patterns 1861-1971 . H.M. Stationery Off.

[2] Lima, C. H. R. & Lall, U. 2009 Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak, and end. Water Resources Research 45 (7).

[3] Liu, Y., Zhang, W., Shao, Y. & Zhang, K. 2011 A comparison of four precipitation distribution models used in daily stochastic models. Advances in Atmospheric Sciences 28 (4), 809–820.

[4] Lunn, D., Jackson, C., Best, N., Thomas, A. & Spiegelhalter, D. 2012 The BUGS Book: A Practical Introduction to Bayesian Analysis. London: Chapman and Hall.
[5] Marin, J.-M., Mengersen, K. & Robert, C. P. 2005 Bayesian modelling and inference on mixtures of distributions. Handbook of Statistics 25, (D. Dey and C. R. Rao, eds.) volume 25 (to appear), 15840–15845.
[6] Masson, P. R. 2001 Exchange rate regime transitions. Journal of Development Economics 64 (2), 571–586.
[7] McCullagh, P. & Nelder, J. A. 1989 Generalized linear models (Second edition). London: Chapman & Hall.
[8] Met Office 2017 Meteorological Office land surface stations observations 2017. http: //data.ceda.ac.uk/badc/ukmo-surface/data/united_kingdom/.
[9] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21 (6), 1087–1092.
[10] Monahan, J. F. 1983 Fully Bayesian analysis of ARMA time series models. Journal of Econometrics 21 (3), 307–331.
[11] Nelder, J. A. & Wedderburn, R. W. M. 1972 Generalized linear models. Journal of the Royal Statistical Society, Series A, General 135, 370–384.
[12] Newcomb, S. 1886 A generalized theory of the combination of observations so as to obtain the best result. American Journal of Mathematics 8 (4), 343.
[13] O'Hagan, A. 1998 Eliciting expert beliefs in substantial practical applications. Journal of the Royal Statistical Society: Series D (The Statistician) 47 (1), 21–35.