The Econometrics Journal

Econometrics Journal (2010), volume **13**, pp. B1–B5.
doi: 10.1111/j.1368-423X.2009.00306.x

BOOK REVIEW:

A Review of *A First Course in Bayesian Statistical Methods*

By HOFF (PETER D.) (New York, NY: Springer Science + Business Media LLC: 2009. Pp. 268. £53.99, hardcover, ISBN: 978-0-387-92299-7)

INTRODUCTION

Recent years have seen an increase of interest in Bayesian statistical methods in many fields, including econometrics. However, many Ph.D. students and empirical researchers have had little or no exposure to Bayesian methods. For these people, this slim volume is the book to read. It offers a clear and concise introduction to the why's and how's of Bayesian statistics. Despite its title (i.e. as a 'first course'), it assumes the reader has a great deal of previous knowledge of probability and statistics. It is not intended as a first course in statistics, but rather a first course in *Bayesian* statistics for the reader already familiar with frequentist statistics. Hence, its intended use lies in advanced level courses in statistics and as a book to read for frequentist statisticians interested in learning more about Bayesian methods.

I divide my review into two parts. The first is a conventional discussion of coverage, strengths and weaknesses of the book. The second is a discussion of the usefulness of this statistics book for an econometrics readership.

COVERAGE

The author adopts a bold strategy: the book begins by addressing the 'Why Bayes?' question after only a one-page introduction of what Bayesian learning is. This strategy works well. By focusing on two examples (the first estimating the probability of a rare event, the second a predictive exercise), some essential features of Bayesian analysis are illustrated and comparisons with frequentist methods are made. The first of these examples makes clear the implications of the Bayesian practice of conditioning on the data. The comparison between frequentist confidence intervals and Bayesian credible intervals is a clever one which draws out the distinction between pre- and post-experimental coverage in a nice way. The second example, involving prediction in a regression with many explanatory variables but only a moderate sample size, is very relevant for modern macroeconomic forecasters who have learnt the value of shrinkage in improving forecast performance. This example compares Bayesian forecasts with OLS-based ones, discusses why the former are superior and relates Bayesian methods to lasso methods. This general pattern, of explaining Bayesian methods in relation to familiar frequentist methods, is adopted throughout the book (e.g. in discussions relating frequentist p-values to Bayesian concepts, examination of the bias and mean squared error of Bayesian estimators, etc.).

The second chapter goes back to the basics: developing the basic tools of probability that form the heart of Bayesian statistics. The flavour of this chapter is somewhat different than a comparable frequentist book. Following common Bayesian practice, it begins with a notion of probability based on a degree of belief. It begins with a standard set of axioms of belief and then proves (in a very clear and succinct fashion) that probability functions satisfy these axioms. The concept of exchangeability is fundamental to Bayesians and, thus, receives appropriate emphasis. The chapter ends with de Finetti's theorem, showing how exchangeable beliefs for that data can be equivalently expressed in terms of a prior and a likelihood. With this result, the reader is now set to see how Bayesian methods work in practice.

Many of the following chapters simply go through various models (e.g. Binomial, Poisson, exponential family, normal, multivariate normal, linear regression, etc.) and, in this review, I will not discuss each model and chapter individually. Instead I will discuss some important themes in modern Bayesian statistics and discuss how the author treats them.

The ability to derive an appropriate posterior simulation algorithm and produce a computer program for implementing it are important skills in Bayesian econometrics. These skills are sometimes unfamiliar to frequentist econometricians. Apart from a few simple models, the posterior and predictive distributions used by the Bayesians do not have analytical forms. Thus, it is important for a book such as this one to provide adequate coverage of posterior simulation and computer programming. To a great extent, the coverage of these topics in this book is of high quality. Early on in the book, the author brings in the idea of posterior simulation, beginning with Monte Carlo methods. The Gibbs sampler, which is used with so many econometric models, appears in Chapter 6. Metropolis–Hastings is covered in Chapters 10 and 11. The algorithms are explained in a simple and intuitive manner, although often full details of proofs are omitted. For instance, the Gibbs sampler converges to the required posterior under certain regularity conditions. The book deals with these regularity conditions by saying 'Under some conditions that will be met for all of the models discussed in this text . . .'. Similarly the author's proof of the Metropolis–Hastings algorithm is labelled a 'proof' (basically the proof is done for only a discrete case). However, this is not a weakness. Such mathematical formalism is not required by the intended reader of this book. Having an informal description of how and why such algorithms work, as provided in this book, is much more useful (I found the section on 'Why does the Metropolis–Hastings algorithm work?' particularly good).

This book is a very concise one, where many important concepts are explained in a very succinct manner. Thus, I was surprised to see the discussion of MCMC diagnostics in Chapter 6 to run to seven pages (and there are additional pages on this topic in later chapters). However, for the econometrician learning MCMC methods for the first time, it is perhaps useful to know this material. The step from a theoretical understanding of how MCMC works to actually doing MCMC in an empirical exercise can be a big one. Knowing how to monitor convergence of an algorithm is an important skill for any empirical Bayesian. The example used by the author in Chapter 6, involving a posterior with three modes, is well chosen to show how you can go wrong when doing MCMC and what you can do to make sure you do not go wrong.

To aid the reader in understanding how the computer programming associated with such algorithms is done, the author provides R computer code as text in the book for many of the empirical examples (even though data and code are also available on the author's website). Because the models used are relatively simple ones, these codes are relatively short (e.g. they tend to cover approximately half a page or a page at most). But for the reader who is unfamiliar with R (such as myself), these can be a bit hard to follow and could do with some more comments. My own preference is for sketches of the structure of the computer code (such as the author includes when e.g. describing the Metropolis–Hasting algorithm) and then putting detailed code on a website. But no doubt R users will find the R code provided directly in the book useful.

Programming is a useful skill for any empirical Bayesian and most Bayesian econometricians create their programs from scratch in languages such as R, MATLAB or Gauss. However, there are an increasing number of Bayesian programs or code repositories that allow the researcher to avoid much of the programming task. The Bayesian program WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/) is enjoying increasing popularity. The R-repository CRAN has many Bayesian packages (e.g. bayesm available at http://cran.r-project.org/web/packages/bayesm). Discussion of resources such as these is not provided in this book.

Posterior simulation and computer programs are two aspects of Bayesian statistics which are often unfamiliar to frequentist econometricians and, hence, a strong feature of this book is its treatment of these issues. Another aspect unfamiliar to frequentists is prior elicitation. A point emphasized in this book is that Bayes' theorem provides an updating rule for combining prior and data information. Priors are not right or wrong, but useful or not useful. A variety of different approaches, including the use of diffuse priors, subjectively choosing prior hyperparameters to match prior expectations, training sample priors, priors based on previous studies, etc., are presented. The usefulness of prior sensitivity analysis is also emphasized. There is also a chapter on hierarchical priors. Hierarchical modelling has played such an important role in modern Bayesian econometrics that it is good to see it getting a decent treatment here.

Econometricians are interested in estimation, model comparison/selection (hypothesis testing) and prediction. This book contains a great deal of material (in terms of theoretical derivations, computation and in the empirical illustrations) about estimation and prediction, but relatively little about model comparison. The concept of a Bayes factor is briefly described at the very beginning of the book and then does not re-appear until Chapter 9 (which involves the linear regression model). Formulas for the marginal likelihood in standard models are typically not provided. Commonly-used tools for calculating Bayes factors (e.g. the Savage–Dickey density ratio) and marginal likelihoods using Gibbs sampler output are not discussed. However, in a short book an author cannot cover everything. There is a substantial discussion of posterior predictive model checking. Although such checks of model fit are not commonly used in Bayesian econometrics, their use and relation with frequentist diagnostic checks are made clear in this book. Various methods related to cross-validation, which involve withholding part of the data and comparing predictions to withheld data, are described.

It is only in Chapter 9, for the linear regression model, that a more lengthy discussion of model selection and averaging is provided. The discussion of model comparison here is elegant, but focuses on a narrow range of issues. The set-up is for the common problem where the researcher is working with a regression model with a large number, K, of potential predictors, many of which are expected to be unimportant. A set of models is defined by introducing $z = (z_1, \ldots, z_K)'$, where $z_j \in \{0, 1\}$ indicates whether a predictor is included or excluded. That is, a conventional regression model is replaced with

$$y_i = \sum_{j=1}^{K} z_j \beta_j x_{ji} + \varepsilon_i.$$

Results for comparing different models defined by different configurations for z are provided assuming a g-prior is used. A posterior simulation algorithm for drawing z is provided (this is required for the common case where z can take on 2^K configurations and K is large). Such algorithms are widely used in contemporary Bayesian econometrics (e.g. in the cross-country growth regression literature where Bayesian model averaging methods have been influential), so such material is very useful. However, the book provides no broader discussion of hypothesis testing in the linear regression model and focuses only on the g-prior. For broader discussion, one must turn to a conventional Bayesian econometrics book.

The final chapters of the book offer a brief introduction to the use of Bayesian methods and posterior simulators in more sophisticated models. Chapter 11 discusses general linear-mixed effects models and Chapter 12 discusses latent variable methods for ordinal data. I found of particular interest in the latter chapter the concept of rank likelihood methods and their use in Gaussian copula models.

HOW USEFUL IS THE BOOK FOR ECONOMETRICIANS?

This book is a slim (approximately 250 pages long) statistics volume. So it is unfair of a reviewer such as myself to judge it as compared to a list of desired features in a more lengthy econometrics volume. Nevertheless, I will be unfair and try and answer the following question: 'How good would this book be for a frequentist Ph.D. econometrician or empirical economist interested in making the leap to the Bayesian econometrics research frontier?'

There are some obvious (and mostly unimportant) issues that distinguish this book from an econometrics one. The examples are mostly non-economic (e.g. examples involve ice core data, diabetes, oxygen uptake, etc.). And some of the terminology differs from that used in Bayesian econometrics (e.g. the terminology marginal likelihood is used in a different sense than the Bayesian econometrician would).

If I go through a standard econometrics textbook, then many standard econometrics topics are not explicitly covered. However, this is not to be expected. A reasonable expectation is that a reader of this book could be prepared to dive into the relevant Bayesian econometrics literature. And, to a large extent, I think this book meets this expectation. For instance, there is a large Bayesian panel data literature in which the focus is on modelling individual heterogeneity (e.g. in slope coefficients in panel data regressions). There is nothing explicitly on this topic in the book, but Chapter 8 discusses hierarchical modelling of grouped data. This chapter describes the concepts (both theoretical and computational) which underlie Bayesian panel data methods. I expect a knowledge of this chapter would prepare the reader well for the Bayesian panel data literature. Similarly, the brief discussion of generalized linear models offers some basic insights into models for qualitative and discrete choice popularly used by econometricians.

Mixtures of various sorts have played a big role in modern Bayesian econometrics (e.g. in allowing for more flexible error distributions or more flexible modelling of conditional means/variances in regression models). There is little on this in this book, but the (very brief) discussion of generalized linear mixed effects models offers an insight into some basic ideas of mixture modelling.

On the other hand, traditional econometric issues such as endogeneity (e.g. instrumental variables methods) are not discussed. Nor are time-series methods and concepts that are popular with econometricians (e.g. unit roots and cointegration).

Bayesian econometrics has been revolutionized by the insight that many of our models can be written in terms of (typically high-dimensional) latent variables involving straightforward MCMC algorithms. Examples include probit, tobit, the stochastic frontier model, state-space models, Markov-switching models, structural beak models, various regime switching or threshold models, random coefficient panel data models, various semi-parametric (or flexible) regression models, stochastic volatility models, etc. This book (with some exceptions, e.g. in Chapter 12) focuses on relatively simple parametric models and, thus, these benefits of the Bayesian methods may not come through clearly to the reader. To give an example, vector autoregressive (VAR) models are an old favourite of the Bayesian macroeconometrician. Through a discussion of multivariate Normal models (and empirical illustrations of the benefits of shrinkage through priors in regression contexts), the book prepares the reader very well for delving into the Bayesian VAR literature. However, the book will offer less of a preparation for the researcher interested in doing Bayesian empirical work with the time-varying parameter VAR (TVP-VAR) model or other non-linear VARs which form the basis of much recent empirical macroeconomic research. I would argue that adding material which would prepare the reader for this literature would be well within the structure and purpose of the present book. For instance, latent state models such as the state-space model, which form the basis of the TVP-VAR or Markov-switching VAR, could have been fitted into this book. The relevant MCMC algorithms are easy to explain and such models are used not only by econometricians, but in many other areas of statistics.

However, I can think of few better books that explain so clearly and succinctly the computational tools used by the Bayesian econometrician. Bayesian inference in most econometric models is carried out using Gibbs sampling or the Metropolis–Hastings algorithm. The reader should come away from this book with a good intuition of how and why these methods work, the kinds of problems that can arise when using them and methods for surmounting these problems.

In short, even though this book does not cover all the models used in econometrics, it does cover most of the ideas and methods (both theoretical and computational) that underlie Bayesian econometrics.

CONCLUSIONS

This is an excellent book for its intended audience: statisticians who wish to learn Bayesian methods. Although designed for a statistics audience, it would also be a good book for econometricians who have been trained in frequentist methods, but wish to learn Bayes. In relatively few pages, it takes the reader through a vast amount of material, beginning with deep issues in statistical methodology such as de Finetti's theorem, through the nitty-gritty of Bayesian computation to sophisticated models such as generalized linear mixed effects models and copulas. And it does so in a simple manner, always drawing parallels and contrasts between Bayesian and frequentist methods, so as to allow the reader to see the similarities and differences with clarity.

GARY KOOP

University of Strathclyde