Chapter 9
Bayesian Econometrics

Rodney Strachan

Abstract This chapter provides a survey of the evolution of techniques used in Bayesian econometrics. Since being first proposed in the 1700s, moving into the fields of probability and statistics and then into econometrics in the 20^{th} century, the Bayesian approach has, in the past 50 years, made contributions in most areas of econometrics. Early approaches relied upon either analytical solutions or deterministic approaches to approximating integrals with, it could be said, limited success. It is possible that the lack of suitable tools limited the utility and broad adoption of the Bayesian approach in econometrics until the late 20^{th} century. It was the arrival of improved computational and algorithmic power that greatly increased the possibilities for Bayesian inference. While computing power (computing speed and memory capacity) and advances in computational algorithms have played an important role in the take-up of Bayesian econometrics, a third influence was developments in model specification. The chapter will identify the Bayesian inferential techniques that have endured and found wide application. It will also discuss those that have fallen into relative disuse, including some of which have later found new purpose and re-adoption. The chapter concludes with a brief overview of broad areas of application in empirical economics and the models and techniques used in those areas.

9.1 Introduction

In econometrics and much less statistics, the Bayesian approach spent much of the twentieth century in the margins of applications and was on occasion the object of intellectual controversy. Viewed with skepticism by some in the early part of the century, it subsequently entered into a period of rapid methodological development and expansion in its application from the 1970s on. The end of the last century saw it become one of the central methodological frameworks in empirical economic

Rodney Strachan
University of Queensland, e-mail: r.strachan@uq.edu.au

analysis. The increase in the adoption of the Bayesian approach in econometrics was aided by techniques, primarily computational algorithms, that leveraged advances in computing speed and memory. Notwithstanding these many recent additions to the Bayesian toolkit, techniques that were first employed in the 18^{th} century still find use today, although often in ways their first proponents could hardly have imagined. This chapter explores the historical development of Bayesian econometrics from its origins through to its current state. The remit for chapters in this volume is to discuss the techniques that have have fallen out of general use and those that continue to be widely applied. This narrative does not fit the story of Bayesian econometrics so well as many techniques do not die, so much, as lie dormant until they are reborn. That is, the narrative is more one of techniques that have been usurped in their primary purpose but have found new applications, usually in combination with new techniques or methods.

There have been many excellent historical reviews of the Bayesian revolution from a range of perspectives, some covering the people who advanced the Bayesian approach, others covering the advances in computational techniques, their adoption in statistics generally and, in some studies, more narrowly into econometrics. It is not possible to mention all of the reviews related to Bayesian econometrics so I point to only a few that have informed this chapter. BaŞtürk, Çakmaklı, Ceyhan and van Dijk (2014) trace the adoption of Bayesian approaches, beginning with early criticism of the likelihood methods used in Cowles Foundation Monographs, they then trace the subsequent publication history of Bayesian econometrics. Although this paper begins with a clear focus on Macroeconometrics and time series, starting with the structural models of the 1950s, they do give attention to developments more broadly. In trying to understand the path that Bayesian econometrics has taken, they use citation analysis to track the publication and citation patterns among the major journals in econometrics from the 1970s until early 2014. They point to clear patterns in journals that publish frequentist and Bayesian papers. This work may prove informative for the young researcher wanting to know where she should submit her latest work. They also identify papers and authors that made substantial contributions to Bayesian econometrics and briefly outline efforts to bring together statisticians and econometricians. An excellent overview of the history of Bayesian computation broadly, that is without a focus on econometrics, is presented in Martin, Frazier and Robert (2024b). This paper identifies major developments over the 250 years spanning the 1770s through to the 2020s. Koop and Korobilis (2010) give an historical overview but with a narrow focus on time series methods in macroeconometrics. A broad, somewhat earlier history is provided in Qin (1996) with what could be called an update, providing excellent references to early Bayesian econometric work, in Zellner (2008).

Useful textbooks for teaching Bayesian econometrics include Koop, Poirier and Tobias (2007) and Chan, Koop, Poirier and Tobias (2020), with careful presentation of techniques and their applications. Similarly, Geweke, Koop and Van Dijk (2011) is useful for graduate students as it covers applications from economics and related disciplines, including macroeconomics, microeconomics, finance, and marketing. In the final section of this chapter, general areas of application and some important models are briefly discussed.

The techniques mentioned in this chapter tend to have wide application and so we give only a few descriptions of their employment in empirical work. There are many techniques developed for problems in specific areas such as finance, macroeconomics, microeconomics and marketing. Only some of these will be given a brief mention to keep the chapter of a manageable length. Finally, an effort is made in writing this chapter to give a roughly chronological overview of the techniques that have been used in Bayesian inference. However, as this is a story of techniques that can apply in many settings, and often these techniques find new use in new settings over time, a strict chronological sequence is not possible.

9.2 The Bayesian Framework

The Bayesian approach aims to quantify uncertainty about an unknown vector of parameters, given a known sample of data. It is assumed that the data are generated from the chosen model, taking into account the researchers' beliefs about the parameter. If there is uncertainty about the specification of the model, which there usually is, then uncertainty about the model can also be accommodated in the analysis. The parameters are related to the data via the likelihood and the beliefs are captured by 'the prior' distribution of the parameters.

The posterior distribution of some function of the parameters is the object of inference and can be accessed from the posterior density of the parameters themselves. This posterior for the parameters can in turn be constructed as the ratio of the likelihood times the prior density for the parameters, divided by the normalizing constant of the posterior (the integral of the likelihood times the prior with respect to the parameters). This normalizing constant is known as the *marginal likelihood* and, importantly, is a measure of the empirical support for the model. From the posterior, we are generally interested in *expectations* of functions of the parameters and, in some cases, such as in forecasts, the data. Common examples of functions of parameters include moments and indicator functions for events, such that the expectation of that indicator is the probability of the event.

As suggested, the marginal likelihood is an important tool in model evaluation. For example, if our model is in a set of models under consideration, and we have a measure (the marginal likelihood) and a prior probability for each of these models, we can compute the *posterior probability* of each model *conditional upon a set of models* as proportional to the product of the marginal likelihood and the prior probability of that model. The conditioning of the model probability on the set of models is important in that we usually assume the set is finite. An alternative to model probabilities, which does not depend upon a finite model set and the prior model probabilities, is to compare the evidence on two models at a time. That is, we compute the *Bayes factor* as the ratio of the marginal likelihoods for two models. When using Bayes factors and considering several models, it is common practice to choose a base model, which is usually suggested by the problem under investigation, and to then report for each model the Bayes factor for each model to the base model.

As discussed above, an aim in Bayesians inference is to estimate the expectation of a function of interest and the evidence on models contained in the marginal likelihood is a feature of the approach that distinguishes it from the frequentist or sampling theory approach. This chapter will consider techniques that have been developed and employed at various times for computing estimates of objects of interest, model probabilities, marginal likelihoods or Bayes factors. For this reason, the very light conceptual introduction presented above will be useful when discussing these techniques, as well as for setting the framework for understanding the Bayesian approach. This chapter will walk through many of the approaches that have been used to obtain these estimates.

9.3 From Inception to Awakening in Econometrics

While there were many people who played a role in forming the ideas and then bringing the Bayesian approach into econometrics, there are a few who stand out for their importance. The Reverend Thomas Bayes passed away in 1761 without publishing his eponymous theorem, or the application to which he was applying it. After finding it among Bayes' papers after his passing, Richard Price first presented Bayes theorem in 1763 in his presentation of Bayes and Price (1763). The approach was rediscovered and termed *inverse probability* by Laplace in 1774 (see Stigler (1986)). While adoption of this approach was somewhat muted for the first 150 years or so, it has made significant contributions to econometrics over the latter half of the 20^{th} century and into the 21^{st} century. Arguments in the 20^{th} century supporting the Bayesian approach generally came from work on logic (de Finetti, 1937), probability (Ramsey, 1964) and frequentist statistics (Stein, 1956). It is difficult to point to the beginning of this revival, having begun first in the areas of statistics and probability before the 1950s and into the 1970s. For a relatively wide discussion of the orgins of Bayesian analysis before being taken up in econometrics, see Fienberg (2006). For an early example exploring the workhorse of econometrics, the linear model, see Lindley and Smith (1972). By whatever path the Bayesian approach made its way into econometric analysis, it is clear that Arnold Zellner, motivated by his reading of Jeffreys (1939), played an important part in its promotion in econometrics (McLure, Turkington & Weber, 2010).

As highlighted in the introduction, a distinguishing and sometimes challenging aspect of inference in modern Bayesian econometrics (or the Bayesian approach in most areas) is the predominant aim of obtaining expectations of objects of interest, such as parameters, losses, and probabilities. This contrasts with the frequentist approach which relies upon obtaining optima such as in maximum likelihood estimation or the various least squares approaches. For a long time, Bayesian estimates for many problems were either inaccessible or very difficult to obtain with any accuracy. This was due to the lack of closed-form expressions to these expectations or the lack of accurate or even achievable deterministic approximations in problems of reasonable dimension. As econometrics has progressed, with an increasing understanding of the importance of data for informing decisions, we have seen an increase in our ability to collect, record and retrieve data, as well as an increase in the dimensions and complexity of models. Without progress in computational techniques, the many high dimensional problems that present in modern models would have proven impossible to investigate accurately or at all. We begin with an overview of these early tools and move onto give a brief outline of the many substantial developments, although not in a strict chronological order, to trace the gains we have seen.

In his very first application of the Bayesian approach, Bayes attempted to evaluate an expectation from the posterior using quadrature. This was only partially successful and may explain why Bayes did not publish his work (Martin et al., 2024b). The approach proposed a few years later, and apparently independently of Bayes' work, was the *Laplace approximation* (Laplace, 1986). This approximation proved far more successful and remains in use today in ways that Laplace could not have envisaged. In his paper, Laplace presented a proof of consistency of the posterior in a specific application and, in developing the proof of this approach, proposed an asymptotically valid approximation to posterior expectations of functions of interest. His approach used a Normal and a second-order Taylor series approximation to the posterior to obtain an estimate of the expectation.

Laplace was attempting to compute the probability in an interval for a Beta posterior. Almost 200 years later, the approach he proposed began to attract further interest in more general settings such as in Tierney and Kadane (1986) and Tierney, Kass and Kadane (1989). This approximation has the advantage of speed and that it relies upon relatively few, generally reasonable, assumptions. While not as widely employed as many of the MCMC approaches, the Laplace approximation has found further new use in a number of applications. In a recent example, which we discuss further in this chapter, Rue, Martino and Chopin (2009) propose the *integrated nested Laplace approximation* or INLA. Work on the same class of approximations continues. Chan et al. (2020), for example, provide a range of Normal approximations to the posterior that rely on assumptions that are met in most applications. They place the Laplace approximation in this set of approximations, however, the original approach continues to be proposed for more general and complex processes (see, for example, Strachan & Inder, 2004 and Arellano &

Bonhomme, 2009 for two very different settings).

One of the oft-stated advantages of the Bayesian approach is that it provides finite sample inference. Having mentioned the Laplace approximation above, however, it is worth pointing out to students interested in pursuing the Bayesian approach to statistical inference, that they should be aware they will not escape the need to have a good understanding of *Asymptotic theory*. This is certainly true of the MCMC methods. Asymptotics have played an important role in achieving inference since Laplace.

Modern interest in the Bayesian approach began to stir in the 1900s. Prior to entering econometrics, however, much work appeared in the statistics literature first, and many of the most successful tools were not originally presented for Bayesian analysis.

Early applications in econometrics relied on analytical results or simple, often one-dimensional, approximations. An early attempt to promote the use of Bayesian analysis in economics was the presentation of conjugacy in Raiffa and Schlaifer (1961) in the context of decision-making in economics. In their book, inference was obtained either from numerical integration with respect to low-dimensional distributions or, where possible, using analytical results. Work around the 1960s and early 1970s tended to use such analytical results or numerical integration, or a mix of both (see, for example, Chetty, 1968a, 1968b, and Drèze, 1976, 1977).

When investigating the linear regression model, econometricians are generally interested in the values of the regression coefficients in the mean. Lindley and Smith (1972) introduced the idea of incorporating prior beliefs into the estimation of the linear Gaussian model. To introduce the algebra involved in Bayesian inference in this model, the variance of the data was initially assumed known such that the estimates, in this case the mode of the posterior distribution of the regression coefficients, could be expressed in a closed form. With advances in prior specifications (such as the inverse Gamma for the variance), this approach has been replaced with the more satisfactory case of unknown variance.

As mentioned earlier, Arnold Zellner played a pivotal role in applying Bayesian methods to econometrics and did much to promote its adoption, especially in regression and simultaneous equations models. Importantly, he generalized from the single equation models to multiple equation models and presented cases in which the marginal posteriors and marginal likelihoods have closed form expressions. Within this class of models, Zellner developed the Seemingly Unrelated Regressions (SUR) model, an important case of which is the vector autoregression model, and contributed to Bayesian model comparison and prediction. His 1971 textbook, Zellner (1971), is still a useful first reference, and valuable introduction to the necessary algebra, for graduate students moving into Bayesian econometrics. Zellner's textbook presents a range of analytical solutions to inference in models in common use in econometrics at the time, with the occasional reference to examples of work in the literature using numerical integration. Numerical integration has largely, but not entirely, disappeared from practice due to both the increasing complexity of models and posteriors, as well as the increasing dimensions of the marginal distributions that need to be explored.

Many of the analytical solutions in Zellner (1971) were possible because of the use of convenient prior forms. However, modern applications require higher-dimensional models and priors with attractive properties (other than producing analytical solutions) that generally result in posteriors with no closed-form solutions to marginal distributions. Fortunately, there have been a number of methodological and technological changes over the past 50 years that have made the Bayesian approach more accessible. An important methodological shift was the introduction of sampling methods using computer-generated random draws to explore the posterior distribution. The major technological shift that has occurred can be described as improvements in computational power. The use of the term 'computational power' here is not intended to simply refer to the increases in computation speed, including speed in accessing memory, or the decrease in the cost of memory. There were also improvements in computational algorithms that were able to, and often designed to, take advantage of this increased speed and memory in producing random draws from the posterior. This chapter will outline a number of important algorithms and the evolution of these computational solutions to evaluating the posterior distribution.

While analytical results are remain useful for improving the efficiency of sampling schemes, such as in the partially collapsed Gibbs sampler (van Dyk & Park, 2008), inferences based entirely on analytical results are rarely available. There has also been a tendency for priors to work more as regularity functions, such as those used in frequentist methods, in that they tend not to necessarily or only represent prior beliefs. Rather, they have been designed to induce good finite sample, or good forecasts, or asymptotic properties in the posteriors such as consistency or smaller expected frequentist risk as in Ni and Sun (2003). Further, the integral of interest may be highly nonlinear (see, for example, Geweke (1988)) and so not amenable to analytical solutions. This has resulted in posteriors that do not lend themselves to analytical integration.

9.4 Early Computational Methods

As Bayesian inference often requires the evaluation of integrals with respect to the posterior distribution that cannot be solved analytically, asymptotic theory and sampling approaches have been used to permit the estimation of such integrals. An idea that opened up opportunities for inference was the application of asymptotic theory, not in the direction of the size of the data sample, but in the number of random draws from a distribution. It was understood for a long time that it was possible to use simulations, such as those produced using a computer, to approximate draws from the posterior distribution. That is, if a sequence of draws 'look' like they come from the distribution of interest, then, if some reasonable assumptions are met, we can draw upon simple asymptotic theory to ensure that the sample average of draws of the function of interest will converge to the expectation of that function under the posterior.

An excellent example of the power of sampling to provide an answer to a complex problem is presented in Geweke (1988). This paper presents a case of using a Bayesian approach that resulted in inference not possible from a frequentist approach. In that paper, Geweke evaluated the empirical support for a " ... *third-order autoregression with a pair of complex roots whose amplitude is smaller than the amplitude of the real root."* Such a restriction would be difficult to evaluate without a sample from the posterior, but is quite straightforward with one.

In the frequentist approach, arguments are made to justify the use of consistent estimators in finite samples. Using the same arguments, Bayesians assume that, given the size of the sample of data, the number of draws from the posterior is 'large enough' to ensure that the error of the finite sample estimator is 'small enough'. Important distinctions between the application of this heuristic in the Bayesian use are that Bayesians aim for inference with finite sample size, but can choose the number of posterior draws, and so have some control over the size of the error. How many draws are taken from the posterior, however, is not necessarily the relevant measure of 'large enough' when the draws are a dependent sequence. If the draws are dependent, such as those drawn from a Markov chain, then this number needs to be adjusted for comparison with other ways of obtaining draws. The standard is to compare to an independent sequence of draws. Bayesians tend to think of the *effective sample size* (ESS) as an appropriate measure (see, for example, Chapter 11 of Gelman et al. (2013)). This is an adjustment to the number of dependent draws to give a measure of the equivalent number of independent draws. The ESS provides a relatively easy-to-understand, and compare, measure of the performance of a sampling algorithm.

There was much literature on Bayesian computation from the 1970s through to the end of the 1990s that was devoted to addressing the problem of efficiently sampling from the posterior. Before we discuss those approaches, we begin with a discussion of estimation without the need to draw from the posterior. One of the first successful sampling approaches employed in Bayesian econometrics was *importance sampling*. This approach was first proposed by Hammersley and Handscomb (1964) for variance reduction, that is, to minimize the standard error of the estimate resulting from a Monte Carlo simulation. The feature of this technique salient to Bayesian computation is that it does not, as mentioned, require draws from the posterior. As many posterior distributions for econometric models are difficult to draw from directly, importance sampling offered the opportunity to obtain Bayesian inference in a wider class of models.

Rather than use draws from the posterior to compute the estimates of interest, importance sampling uses a candidate distribution. Essentially, draws from the candidate are re-weighted to match draws from the actual posterior. That is, the expectation of a function of interest could be evaluated by taking advantage of a simple algebraic identity such that, with a sequence of draws from the candidate, we could estimate this expectation of the function by its weighted average where the weights are, quite simply, the ratio of the posterior density over the candidate density. While this estimator is an important variance reduction technique (Kroese, Taimre and Botev (2011)), it is only useful when we know the full form of the posterior density.

The requirements that must be met to employ importance sampling point to just some of the challenges of this approach. First, estimation will be better if the candidate distribution is chosen such that 'looks like' the posterior. Second, it should be easy to obtain draws from this candidate. The tension between the first two points are evident. Further, it is important to carefully choose the candidate density such that it dominates both the posterior and the product of the posterior and the function of interest (for which we seek the expectation). This is to ensure that weights are stable which in turn ensures stability, in terms of having finite variance, of the estimator.

Another difficulty with the implementation of the importance sampler discussed above is that we need to know the full posterior density. It is more common that we only know the posterior up to the normalising constant or marginal likelihood. That is, we generally know the likelihood and the prior for the parameters, but are not able to evaluate the integral of the product of the likelihood and prior (the marginal likelihood). There is a simple solution to enable importance sampling when the marginal likelihood is unknown that involves a slight adjustment. That is, we divide the sample average of weighted draws of the function by the sample average of the weights. The presence of the weight in the ratio removes the need to know the full, normalized, posterior since the normalizing constant (the the marginal likelihood) cancels.

The paper Kloek and van Dijk (1978) introduced importance sampling to Bayesian econometricians and, in an early application, Bauwens (1984) applied it to inference in the simultaneous equations model. This approach became an important early computational tool for Bayesian econometricians. The importance of Kloek and van Dijk (1978) was that it first taught Bayesian econometricians of the virtues of using sampling of parameters to obtain inference.

Kroese et al. (2011) give a useful discussion of the importance sampler, as well as discussion of the importance of finding the best (minimum variance) choice of candidate density. This leads into a discussion of how to minimize the variance of the estimator, and from there into the topic of cross-entropy (Kullback, 1959) which, roughly speaking, involves concerns around measuring the difference between distributions. Coming from information theory, cross-entropy has a number of uses relevant to Bayesian econometrics such as to obtain candidate densities or marginal likelihood estimation. Cross-entropy demonstrates attractive properties as shown in Chan, Glynn and Kroese (2011) (and further developed in Chan & Kroese, 2012). It has also proven useful in a number of applications such as in nonlinear Gaussian state space modelling. Example applications include Chan and Strachan (2012), Chan, Koop and Potter (2013), Chan and Strachan (2014) and Chan, Koop and Potter (2016). However, cross-entropy has not yet been widely adopted as a related approach that we will discuss later, variational Bayes.

While the importance sampler has the convenient feature of not requiring draws from the posterior, it does require careful consideration of the candidate and the weights to ensure existence of estimates. Using draws directly from the posterior avoids the need to consider candidate densities and the behaviour of weights. Many approaches have been proposed to use samples from the posterior to obtain inference. A problem that existed in the early stages of Bayesian computation, however, and still exists today is that it is often (usually) difficult or even impossible to sample from the full, marginal or conditional distributions. Several early sampling methods faced restrictions that meant Bayesian inference in even moderately complex economic models was not possible. Consider, for example, the *Inverse Transform Sampling*, which shares features similar to a technique labelled the *Griddy-Gibbs* in Ritter and Tanner (1992). Both use the result that, for a random variable with a given distribution, then the distribution will be uniformly distributed over the unit interval. An implication of this is that the inverse of the distribution (as a function of the uniform random variable) will have the same distribution as the original variable. The inverse transform sampler uses this result directly whereas the Griddy Gibbs is useful where the distribution, and so the inverse distribution, are known only up to the normalizing constant. The clear limitation of these approaches is that they are only applicable for scalars.

An early method that could, theoretically, produce draws from the full posterior employed by Von Neumann was rejection, or rejection-acceptance, or accept-reject (AR) algorithms. The AR method continues to find use in a wide range of circumstances, although is rarely used as the primary sampler. Rather it tends to be used as an enabling step in more generally employed samplers such as the Gibbs sampler, similarly to how a Metropolis-Hastings (MH) step is often inserted into a Gibbs sampler (against the advice of Chib & Greenberg, 1995). Indeed, Chib and Greenberg (1995) point out that the rejection sampler shares similar features to the Metropolis-Hastings. For example, the AR sampler uses draws from a candidate density that are accepted with an appropriate probability to obtain independent draws from the target density. Unlike the Metropolis-Hastings algorithm, however, this approach can produce a non-Markovian, usually independent, sequence of draws from the posterior.

The acceptance-rejection (AR) sampler has a relatively simple structure. A candidate draw of the parameter is obtained from a candidate distribution where the candidate density times a constant (greater than one), is greater than the posterior. This draw is accepted with probability equal to the ratio of the posterior density over the constant times the candidate density. This outline of the algorithm helps point to the well-understood limitations to this approach that have restricted its application. The first issue is finding a value for 'the constant'. Ideally this value would be as small as possible since the acceptance rate is proportional to the inverse of the constant. However, this requires finding the supremum of the ratio of the posterior to the candidate density, which is rarely a simple exercise, even in only moderate dimensions. The second issue is the choice of the candidate density. Ideally, this would resemble the posterior in where it places mass. Again, this is not always easy to obtain.

A wide array of adaptations of the AR algorithm have been developed over time to better meet the needs of specific situations or to overcome the issues mentioned. More recently, it has found extensive use when combined with other algorithms as in the accept-reject Metropolis-Hastings algorithm (ARMH) (Tierney, 1994, and Chib & Greenberg, 1995). The ARMH has proven useful in drawing from complicated densities. For example, Chan et al. (2013) propose a model of bounded trend inflation using a state space model. In this paper, the state variable – trend inflation – has bounded support due to the imposition of inflation targets by the central bank. This specification is extended to bound the NAIRU in Chan et al. (2016), again resulting in a nonstandard posterior.

It is possible to improve speed and efficiency of computation in the ARMH algorithm by using a so-called *squeeze*. This involves bounding the target density with a (often piecewise linear) function and evaluating whether a draw from the candidate meets this bound before further considering whether to accept (or reject) the draw. Gilks and Wild (1992) present an adaptive acceptance algorithm in which both the rejection envelope and the squeeze function converge upon the target density as the sampling advances. While this approach has found considerable use in statistics, there are fewer applications in econometrics, although some do exist. Examples include sampling inefficiency factors in stochastic frontiers in Griffin and Steel (2008). Kalli, Walker and Damien (2013) employ it in a nonparametric study of financial returns, and Loddo, Ni and Sun (2011) in a stochastic model selection approach.

Such approaches may be difficult to use in higher-dimensional settings. It is only in a few high-dimensional cases in the literature that we are able to obtain direct draws of the parameter vector from the full posterior. In such situations, we might, instead, be able to take advantage of the structure of the posterior to obtain draws. One such case results in the method of composition which produces direct and independent samples from the posterior (rather than from a candidate). Partition the vector of parameters as two sets of parameters, then decompose the posterior into the marginal distribution for one set of parameters and the conditional distribution for the second set given the first. If we are able to directly draw the first set of parameters from its marginal distribution and then draw the second set directly from its conditional distribution conditional (i.e., conditional upon the first set), we will then have a direct draw of all parameters from the full joint posterior. This approach requires, of course, being able to draw from each of these marginal and conditional distributions. A simple example where this situation arises is the general linear regression model with a Jeffreys prior as in Chapter 8 of Zellner (1971). In these cases, the choice of prior can imply that the covariance matrix will have an inverted-Wishart marginal posterior, while the regression coefficients will have a posterior conditional upon the covariance matrix, that is multivariate Normal. As it is straightforward to draw from these two distributions, we can implement the method of composition to obtain a direct sample from the full posterior.

While the idea of using draws from a distribution to estimate expectations of functions of the variable was long understood, it was some time before it was applied to Bayesian analysis. The challenge in this approach was to obtain draws from the posterior distribution, particularly when the parameter space is of (even moderately) high dimensions and when the normalizing constant of the posterior is not known. This is often the case. A very early general solution to obtaining draws from such 'difficult' distributions was proposed in Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953). The approach was called, initially, the *Metropolis Monte Carlo Algorithm*, or Metropolis Algorithm.

As with importance sampling, the Metropolis algorithm used draws from a candidate density, but rather than re-weighting the function in the expectation, the draw was accepted with probability that depended upon how much closer was the new draw than the last to the highest region of probability under the posterior. The original method produced a candidate at each iteration that was equal to the current draw plus a uniform random variable centred at zero, and accepted with probability (the acceptance probability) that is the minimum of one and the ratio of the posterior evaluated at the proposed draw divided by the posterior evaluated at the previous draw.

An attractive feature of this approach is that the computation of the acceptance probability does not require full knowledge of the posterior. That is, we need only know the posterior of the parameter up to the normalizing constant. This is an advantage as this normalizing constant is often unknown or has no closed form expression. The sequence of draws obtained using the Metropolis Algorithm, unlike earlier sampling methods, was a dependent sequence, specifically a Markov Chain, from the posterior. As the algorithm produced a Markov Chain, the emcompassing methodology has been termed *Markov Chain Monte Carlo*, or MCMC. The introduction of the MCMC approach to obtaining draws from the posterior distribution had an accelerating effect on the adoption of Bayesian econometrics (Geweke (1989)) leading, via the Metropolis-Hastings method, eventually, to the particularly powerful special case in the Gibbs sampler. Seminal papers in this area include Geman and Geman (1984), Tanner and Wong (1987) and Gelfand and Smith (1990). There are many excellent review articles on MCMC methods. For example, some useful reviews can be found in Besag and Green (1993), A. Smith and Roberts (1993), Geyer (2011) and Chib (2011), among many others. For an overview of the early history of MCMC see Robert and Casella (2011).

The Metropolis algorithm was generalized by Hastings (1970) to permit draws of candidates from a wider range of candidate densities. The power of this idea was that it could provide an estimate of the expectation of interest using draws from a distribution that was difficult to sample. This technique found extensive use in physics (Chib & Greenberg, 1995) before it was later adopted into statistics and applied to Bayesian computations generally, and eventually into econometrics.

The introduction of the Metropolis-Hastings algorithm was an important step in the enabling Bayesian econometrics. However, the acceptance rate in the Metropolis and the Metropolis-Hastings algorithms depended upon the choice of a constant in the former and of the candidate density in the latter. These choices can have significant impacts upon the acceptance rate of draws and, thus, the speed at which the sequence converges to the invariant distribution. These choices necessarily require some degree of calibration and often experimentation to find choices that produce an acceptable sequence of draws. The *Gibbs sampler*, presented in Geman and Geman (1984), was a much simpler, if not as broadly applicable, technique that has proven incredibly powerful in enabling inference. Although the Gibbs sampler is a special case of the Metropolis-Hastings algorithm, its simplicity is this sampler's most useful feature. The term Gibbs sampler was coined because Geman and Geman (1984), who applied the sampler to the problem of image reconstruction (Ritter & Tanner, 1992), were working with a posterior that had a Gibbsian distribution. The same algorithm was presented in Tanner and Wong (1987) in their paper proposing Data Augmentation, again, because this often simplifies computation.

The Gibbs sampler is very well known and one of the first samplers graduates learn. We therefore give only it a brief overview. Partition the parameters into two sets. It is often the case that draws of neither of these two sets of parameters is possible from their marginal (that is, marginal of the other set of parameters). But it may be that we are able to draw from the conditionals of each set of parameters, conditional upon the other set. In this case, sequences of draws from these conditionals can produce a sequence of draws whose convergent distribution is the full joint posterior distribution. For an excellent and accessible explanation of this result for the Gibbs sampler, and for the Metropolis-Hastings algorithm more generally, we direct the reader to Chib and Greenberg (1995).

We mentioned earlier the inverse-transform method of drawing from a target distribution. The *Griddy Gibbs* of Ritter and Tanner (1992) uses a related approach in that a histogram approximation to the conditional distribution of a scalar parameter is constructed using a grid of values for the parameter. The inverse transform method is then used to draw the nearest value on the grid as the current draw. This process is repeated at the step to draw this parameter in each iteration. While this is not as elegant as draws from a continuous support, and it is sometimes computationally more demanding, it does enable draws from a scalar distribution that is difficult to access. There remain questions with how fine to make the grid but this technique continues to find use in a range of applications in econometrics, such as Bauwens and Lubrano (1998), Ausín and Galeano (2007) and Droumaguet, Warne and Woźniak (2017).

In a very promising approach to sampling difficult posteriors using Hamiltonian dynamics, Duane, Kennedy, Pendleton and Roweth (1987) introduced the Hamiltonian MCMC, or HMC, method as a Metropolis-Hastings algorithm. Neal (2011) provides a nice overview of the journey of HMC from the study of motions of molecules to Bayesian inference, as well as a clear introduction to its implementation. There have been a number of interesting extensions to the original HMC such as Nishimura, Dunson and Lu (2020) development of an HMC for discrete supports, and the approach to sampling when the support is a d-dimensional manifold presented in Byrne and Giralomi (2013).

More recently, HMC is seeing increasing numbers of applications in econometrics. Already, it has been used in macro and financial applications. Choi, Jo and Kim (2025) use HMC to estimate a time-varying parameter model of the Cobb-Douglas production function. To estimate their multivariate financial time series model with a single-factor copular model and stochastic volatility, Kreuzer and Czado (2021) overcome the dimension limitations encountered with existing approaches by developing an HMC sampler. In estimating a singular vector autoregression model for a structural analysis of the US macroeconomy, Eisenstat and Strachan (2026) use HMC to draw from both discrete mass functions for the VAR lag length and the rank of the system using the approach in Nishimura et al. (2020). They also use the extension of Byrne and Giralomi (2013) to draw from a Steifel manifold.

An important conceptual advance in the use of MCMC methods was the inclusion of latent states (parameters) ("Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables" (1999)). An example of this is the *slice sampler* of Neal (2003). This approach augments the parameter vector with a latent state, or auxiliary variable, from a uniform distribution to draw from posteriors of parameters that may be multidimensional. Chapter 8 of Robert and Casella (2005) provides a good introduction to this tool and Herce and Salvador (2026) outline interesting extensions.

In another approach using latent variables, *data augmentation* (DA) has been used to obtain inference in many otherwise inaccessible problems. This approach is particularly useful when, and was first proposed for, applying the EM algorithm for maximum likelihood estimation (Dempster, Laird & Rubin, 1977). Tanner and Wong (1987) presented it as a solution to problems with missing data. More importantly, they show how this technique can be applied for computation of posterior distributions by taking advantage of the case where the likelihood or posterior distribution are simplified when considered conditional upon augmented data.

In DA, the posterior is augmented with latent data with a predictive density dependent upon the observed data. This could be useful in cases where it is difficult to calculate the posterior, but we can construct an 'augmented' posterior, that is augmented by some latent variable z, which is relatively easy to compute. If we are able to augment the observed data with latent data to obtain augmented data in such a way that we can draw the latent variable conditional upon the data (and marginal of the parameters), then we can draw the parameters from their posterior conditional upon only the observations by integrating out the latent variables. This marginalization can be achieved by Rao-Blackwellizing the posterior of the parameters condition upon the observed and latent variables with respect to the latent variable. It is worth noting that the components of the DA sampler in Tanner and Wong (1987) were the same as the well-known Gibbs sampler, but recognition for introducing this has largely gone to Gelfand and Smith (1990). For careful textbook treatment of the DA approach, see Chapter 5 of Tanner (1993).

The method of data augmentation can be used not only for simulating missing values of data, but it can also be adapted to improve mixing and convergence of the Markov chain for posteriors. One such approach is termed *parameter expansion* (for reasons that will become obvious), and resembles data augmentation in that the posterior is augmented, although with non-identified parameters rather than data. The approach takes advantage of the structure of the posterior and mapping to new parameter structures. Using a judiciously chosen non-identified model may, according to Liu and Wu (1999), Meng and van Dyk (1999) and MacEachern (2007), result in improved convergence of the MCMC chain. One application is the parameter-expanded Gibbs sampling (PX-GS) algorithm for which improvements in convergence and mixing can be achieved using ideas from Liu 1994 and Liu and Wu 1999. See, for example, the application to permit efficient estimation of the Multinomial Probit in Lawrence, Liu, Bingham and Nair (2008), and in the static factor model in Chan, León-González and Strachan (2018).

Going in the opposite direction to data augmentation, Liu, Wong and Kong (1994) and Liu (1994) show that integrating out some parameters, or collapsing the parameters, can provide significant improvements in efficiency and speed. This approach skips draws of the parameters that have been collapsed out as in the *Partially Collapsed Gibbs sampler* (or *Collapsed* Gibbs sampler) presented by van Dyk and Park (2008). More specifically, if we have parameters a set of three parameters, but we are able to draw from the conditional posterior of the first parameter, conditional on the second parameter but marginalised of the third, and we can similarly draw the second parameter conditional upon on the first, we can draw these before going on to draw the third parameter conditional upon the first and second. That the draws of the first two parameters are taken marginal of the third is a simple example of how the chain is *collapsed*.

An important property for an MCMC algorithm to have is that it converges to the target distribution. If the Markov chain we are using has not yet converged, or is not going to converge to the target distribution, the inference we obtain can lead to inaccurate conclusions. A challenge arises in that, as we sample these distributions because we do not know what they look like, convergence of the sample to the target is difficult to determine. As such, there as been much work invested in developing means of evaluating whether a Markov chain has converged. Consideration of such issues come under the broad topic of *diagnostic checking*. We have already discussed diagnotic checking when discussing the measure of effective sample size in which we ask how close is the sequence of MCMC draws to one that is a set of independent draws.

The topic of diagnostic checking has been extensively considered and we cannot mention even a representative sample of the literature. Instead, we point to a few articles that summarize the diagnostic tools. These include Cowles and Carlin (1996), Cowles and Rosenthal (1998), Fan, Brooks and Gelman (2006), Chapter 4 of Geweke (2005) and Chapter 5.4 of Gamerman and Lopes (2006). Accessible overviews of many of the main techniques are provided in Gilks, Richardson and Spiegelhalter (1996) and Roy (2020).

We briefly describe a few prominent examples of measures of convergence. As a measure of accuracy of an estimate, the researcher can report the confidence interval for the estimator. To compute the confidence interval, a standard error of the estimate is required. As the sequence used to compute the standard error is correlated, Geweke (1992) suggests an estimator of the standard error based upon the spectral density. Making further use of this estimate of the standard error, Geweke (1992) proposed another measure of convergence that has a standard normal distribution. This measure is based upon splitting the sample of draws and comparing the early draws with draws nearer the end of the run. This provides a very simple measure of convergence. A relatively recent contribution is Fan et al. (2006) who use the observation that the score of the target distribution is expected to be zero to provide a measure of convergence for a vector of parameters. A measure far from zero, therefore, indicates a lack of convergence.

If the Markov chain has not converged to the posterior distribution but has, instead, become stuck in a region of high probability, it will not provide a good replication of the full posterior. However, the above diagnostics may not pick up that this has occurred. As an alternative to using just one sequence of draws, the researcher could use multiple sequences of draws with different starting points, to evaluate whether convergence has occurred. Gelman and Rubin (1992) propose an early approach to assessing convergence based upon the ratio of total variance to variance within each sequence as measured by $\hat{R}$ (which, by construction, will be greater than or equal to one). It is expected that $\hat{R}$ will decline to one as the sampling continues, where a value of $\hat{R}$ near one is taken as evidence of convergence. Note that this approach does not guarantee that we have achieved convergence, since all chains could become stuck in the same area. But it will likely increase the chances of detecting non-convergence.

A central object needed for methods of obtaining Bayesian inference in all of the approaches that we have discussed to this point, is the likelihood. Situations exist in econometrics, however, where the likelihood is unknown, unknowable, or has no closed form expression. This issue was recognized and a solution proposed using statistics in Zellner, Hill and Fomby (1997) (see also Zellner, Tobias & Ryu, 1998). This introduced the Bayesian equivalent of the method of moments, or Bayesian method of moments (BMoM). This is not a technique that has seen widespread adoption, although related approaches are showing more promise.

Another issue that can constrain full specification or exploration of the posterior arises when the dimension of the model or the data set (‘Big data’) becomes immense such that computation cannot be achieved in a reasonable time. This was certainly true earlier when computing power and memory more constrained. As memory and computing speed increased, many previously inaccessible models could be estimated. However, model size and complexity have continued to increase, as has the size of the data sets. Some data sets now exceed our computational limits. As such, the constraints on computation have re-emerged and we have re-visited some of the earlier techniques to address these limitations. One response to such situations is to adopt approximate solutions, many of which have been developed and improved upon more recently.

While there has not been substantial use or development of BMoM, Martin, Frazier and Robert (2024a) discuss a range of methods, some of which share with BMoM that they are built around using statistics rather than the whole sample. The methods they consider, however, are showing more promise than BMoM in that they are currently being more widely employed and there is considerable work on their development. They provide an excellent taxonomy of the methods by, among other features, their strengths and so give some guidance on appropriate areas of application. The primary approximate methods they identify are Approximate Bayes Computation (ABC), Bayesian synthetic likelihood (BSL), variational Bayes (VB) and integrated nested Laplace approximation (INLA). These techniques assume the posterior is not available, and offer different approximations to posterior inference.

The VB and INLA techniques both require access to the likelihood for the underlying process, but rely on optimization to produce inference rather than a description of the full posterior that we get from MCMC samples. ABC and BSL, on the other hand, do rely on simulation but not of the parameters from the posterior. Instead, ABC draws parameters from the prior (which does restrict the form of the prior that can be considered) and then, conditional upon these parameters, simulates the data. A vector of statistics (which could be, for example, summary statistics or the raw vector of data) constructed from the simulated data are compared to the same vector of statistics from the observed data to obtain an inferred (but not direct) measure of 'closeness' of the parameters to those of the underlying process. The measure of closeness is based upon some (small) level of tolerance. BSL has a similar objective as ABC, being to produce an approximate posterior based upon statistics, but BSL makes use of a Gaussian approximation (for the distribution of the vector of statistics) to the likelihood and does directly consider the distance between the simulated and observed vectors of statistics.

The two approaches ABC and BSL use simulation to approximate the posterior. By contrast, VB and INLA use optimization to obtain the approximation. VB aims to approximate the posterior by minimizing the Kullback–Leibler (KL) divergence from the approximating to the true posterior. This has the advantage over simulation based methods, in certain situations, of speed as it can rely on numerical algorithms. Rue et al. (2009) proposed INLA in which the Laplace approximation is combined with numerical integration to provide a fast approximate method to Bayesian inference. INLA is therefore a deterministic approach to obtaining approximations to marginal posteriors and expectations. By using a mix of approximations of (conditional) densities and numerical techniques to speed computation of operations with high dimensional matrices and numerical integration of hyperparameters, this approach can obtain accurate approximations of the posterior densities of interest. The advantage of this approach is both the accuracy and the speed, relative to MCMC methods for similarly large models.

At this point we diverge from discussing only estimation methods to discuss a class of models and the methods for estimating these models. The linear Gaussian state space model has found extensive use in econometrics, often from a Bayesian perspective. A very useful textbook on the state space model is Durbin and Koopman (2001), although for a Bayesian approach, refer to West and Harrison (1997). A particularly successful seminal application in macroeconometrics can be found in Primiceri (2005). Most Bayesian applications of the state space model use the Kalman filter and smoother to produce estimates. These two estimators have very different interpretations. The Kalman filter produces estimates from a posterior built using only data up to the point in the sample at which the state occurs. We may interpret these as the expectations that an economic agent would have given the information available at that time. By contrast, the Kalman smoother produces estimates using the all information available under the posterior. As with any other estimate of an expectation under the posterior, these Kalman smoother outputs are what we would normally regard as posterior estimates of the states. For technical details on this distinction, see Durbin and Koopman (2001), and for a discussion of the different roles of inference from the Kalman filter and Kalman smoother in Bayesian analysis, see Sims (2001).

A valid criticism of many time-varying parameter models is that they are specified as linear Gaussian state space models; the point of the critique being that they are *linear* and *Gaussian*. Recognition of these limitations and the many processes not captured by such a model has led to new approaches, some Bayesian. An alternative Bayesian approach to the extended Kalman filter for estimating non-linear and non-Guassian state space models was proposed by Gordon, Salmond and Smith (1993). This early *sequential Monte Carlo* (SMC) approach, previously know as the *particle filter* (Gordon et al., 1993, and Kitagawa, 1996), used bootstrapping to obtain a density for the state. The approach gave new life to importance sampling as a way to generate from the very nonstandard distribution of the states (Handschin and Mayne (1969)). As this approach will fail as the sample size increases, Gordon et al. (1993) introduced a second selection step to readjust the importance weights. This set of techniques allowed the range of state space models in econometrics to expand to include a range of volatility models as discussed in Creal (2012) and An and Schorfheide (2007) point to use in nonlinear DSGE models. An excellent resource on SMC is the webpage built by Arnaud Doucet, Doucet (2026).

Unless one is willing to accept some bias in the estimates (Rebeschini and van Handel (2015)), SMC is restricted in the dimension it can handle. Fortunately, alternative approaches exist. Some of these were mentioned earlier as they employ cross-entropy. These are Chan and Strachan (2012), Chan et al. (2013), Chan and Strachan (2014) and Chan et al. (2016). This line of work couples cross-entropy with a second order Gaussian approximation and the precision sampler of Chan and Jeliazkov (2009) to produce estimates of moderately high-dimensional, possibly non-linear, possibly non-Gaussian state space models in a range of applications.

9.5 Model Measurement and Comparison (Including Variable Selection)

Model comparison and model selection are fundamental aims in econometrics and the marginal likelihood and the Bayes factor play central roles in this function (Kass & Raftery, 1995, Geweke, 2005, and Congdon, 2007). These concepts capture evidence on the full range of questions, from the simplest such as the veracity of a linear restriction, to comparing the evidence in support of alternative economic theories. An attractive feature of the Bayesian approach is that it permits comparison of models that do not nest. That is, there is no need for an encompassing model[1]. This

[1] Although, as we will see, the existence of an encompassing model can prove useful for comparing models.

is because the marginal likelihood provides a measure of each model independent of the measure for other models. The posterior model probability is proportional to the marginal likelihood (the evidence in the data on the model) and the prior model probability, which reflects the researcher's relative belief in the veracity of the model.

In the Bayesian approach to model comparison, the marginal likelihood is useful for computing the posterior model probability or for direct comparison of two models via the Bayes factor. The posterior odds of one model to another model is the posterior odds ratio, literally the ratio of the probabilities for the models. If the models being compared are regarded as *a priori* equally likely, such that they have equal prior probabilities, then the posterior odds ratio is the Bayes factor. More generally, the posterior odds ratio is the product of the Bayes factor times the prior odds ratio. Various techniques for estimating functions of the marginal likelihood, Bayes factor or posterior probabilities have been proposed for different settings. Again, the literature on this topic is extensive (see, for just a few examples, Dickey & Lientz, 1970, Gûnel and Dickey (1974), Kass & Raftery, 1995, Verdinelli & Wasserman, 1995, Chib, 1995, Frühwirth-Schnatter, 1999, Ardia, Baştürk, Hoogerheide & van Dijk, 2012, Frühwirth-Schnatter, 2019, and Y. Li, Mallick, Wang, Yu & Zeng, 2025). Llorente, Martino, Delgado and López-Santiago (2023) provide a relatively recent review of these techniques.

In simpler cases, the marginal likelihood can be calculated analytically. For example, analytical results are presented Zellner (1971) for a range of econometric models that were important at the time, but these are mostly based around the linear Gaussian model. A more recent application, also based upon the linear Gaussian model, is presented in Fernández, Ley and Steel (2001a). In this paper, they authors use a *g-prior*, as proposed by Zellner (1986). However, they do not rely entirely upon these analytical results since the model set is too large to compute results for all models. Instead, the approach is combined with the Markov chain Monte Carlo model composition (MC^3) introduced in Madigan, York and Allard (1995). The MC^3 algorithm draws models by taking steps from the current set of included regressors, to a draw from a set that includes the current model and all models that add one or subtract one regressor. Analytical results produce accurate and fast inference. However, for more complicated models or non-conjugate priors, approximations to the marginal likelihood are required and we consider a sample here.

Since analytical solutions do not exist in many situations, we can resort to approximations. As discussed, an early approximation to marginal densities which can be extended to the marginal likelihood, is the Laplace approximation. This approximation comes with an error since it is a truncation of a series (see Ogden, 2021 and Shun & McCullagh, 1995). Given its simple structure and that it uses readily available objects such as the posterior mode and the Hessian of the log posterior, improvements continue to be suggested (e.g., Han & Lee, 2024).

When reporting parameter estimates it is prefered that these are not reported conditional upon other parameter values given these are generally not known. Rather, the preference is to report estimates marginal of all other parameters. That is, we integrate out other sources of uncertainty. It is often the case that interest lies not in the restriction or even the model, but rather some outcome (such as a forecast or impulse response) or other object that may be a function of the parameters (such as an elasticity or a forecast). In this case, it is often preferable to avoid conditioning on any one model so as to reduce specification bias and improve the external validity of the inference. That is, just as we integrate out all parameters rather than condition upon them, we would also like to *average out* the models. This naturally leads to the concept of *Bayesian model averaging* in which the reported entity is the average of outcomes from all considered models, weighted by the model's posterior probability. In such cases where we wish to account for model uncertainty, the posterior probability of a model being true is an important object of interest and there are many ways to compute or estimate this.

Information criteria (IC) provide a useful measure of the model, or its distance from the true model where the distance measure has an information-theortic basis. These IC are often reported in log form and can be related to the remaining entropy (or uncertainty) left after estimating the model. Hence, the closer is the model to the true model, the smaller or lower is the estimated IC measure. Pedogogically, we present these criteria as made up of two components – one that captures how well the model fits the data, and a second that induces a penalty for complexity. The overall measure, therefore, trades off fit and parsimony. Extensively used in time series analysis, for example, is the Schwartz-Bayes information criteria or BIC proposed by Schwarz (1978) as a first-order asymptotic approximation to the (negative of the log of the) marginal likelihood (also discussed in Kass & Raftery, 1995). There have been a number of information criteria with Bayesian justifications proposed. For example, among much interest in bringing the Bayesian approach into time series analysis, Phillips and Ploberger (1994) and Phillips (1995) developed the PIC, or posterior information criteria, specifically for time series analysis.

An important function of econometric analysis is providing advice for policy action or decision-making generally, and perhaps the most commonly employed output for this function is the forecast of economic variables. In a brief comment, Geweke (2001) builds the argument from economic theory and the well-known dictum of Friedman (1953) about the importance of an economic theory's predictive power, to support an argument for using Bayes factors and predictive likelihoods for model evaluation. He shows that the marginal likelihood for a model can be interpreted as the sequence of (in-sample) predictions for that chosen model, and the Bayes factor is a relative measure of predictive performance. Rather than relying on a single model, combining forecasts have been shown to perform well in a number of settings (Petropoulos et al., 2022). Bayesian model averaging provides a natural and theoretically sound framework for combining forecasts and this has been applied to a number of problems (see, for example, Wright, 2008). For further reading, Steel

(2011) discusses the role of variable selection, as a function of Bayesian model averaging, in forecasting and a subsection of van Dijk (2024) outlines the importance of microeconomic foundations to macroeconomic relations in forecasting.

Despite its mathematical elegance and theoretical appeal, Bayesian model averaging has not become a default practice in inference, even when interest is not in the models themselves but their outputs. Asymptotically, one model will have posterior probability one in most circumstances which provides a limiting argument for considering only the most preferred model. In finite samples, however, many models may be informative on the object of interest. In considering forecasts, specifically predictive densities, Geweke and Amisano (2011) promote the use of log predictive scores in optimal prediction pools (of models). An attractive property of this approach is that, even asymptotically, it is possible (and likely unless the true DGP is in the model set) more than one model will have positive weight and so contribute to the predictive density.

Another criterion developed for model comparison or selection when the models have a high dimensional latent component, is the Deviance Information Criteria, or DIC. The DIC, introduced in Spiegelhalter, Best, Carlin and Van Der Linde (2002) with further discussion in Spiegelhalter, Best, Carlin and van der Linde (2014), has advantages over both the BIC and Bayes factors in that, unlike the BIC, the DIC takes into account prior information. It has an advantage over Bayes factors in that DIC can be used with improper priors as it does not suffer from the Jeffreys-Lindley-Barlett paradox. The DIC has seen increasing use, possibly due to its suitability to complex models such as state space models and hierarchical models. Initially missing from the literature, Y. Li et al. (2025) have recenlty provided a decision-theoretic justification for the DIC.

For Bayesian model averaging, the weight for each model can be computed before weighting the outputs (such as parameters or forecasts). An alternative approach is to sample not only the parameters from each model, but also to sample the models themselves. We have already mentioned MC^3 as one approach to doing this. Another algorithm is the Reversible jump algorithm of P. J. Green (1995). The reversible jump MCMC (RJMCMC) algorithm allows to draw from models that potentially have different dimension parameter vectors. Another attractive feature of the RCMCMC is that, as it uses an MH step to jump between models, it does not require a closed form expression for the marginal likelihood. For an extensive coverage and discussion of model averaging, the following selection of articles is useful: Raftery, Madigan and Hoeting (1997), Wright (2008), Lamnisos, Griffin and Steel (2009) and, for recent coverage of work in economics, see Steel (2020).

It is common in econometrics to question the inclusion of various explanatory variables in a model, where exclusion implies the coefficient for that variable is zero. The topic of *variable selection* is a natural area of application for model comparison or model averaging and has attracted much attention generally. In particular, this topic received extensive focus in the Bayesian literature, primarily in the late 1990s and early 2000s. A common Bayesian approach to this question is to report the probability that the variable is included in the model, or the inclusion probability. This question is related to model selection since the restriction induces a new, albeit nested, model. However, when there are many variables, it is usual to report only the marginal inclusion probability for each variable. Without the joint probabilities of inclusion of all combinations of variables, it is not clear how to map to model probabilities. Given the discussion above, it is worth mentioning here Sala-i-Martin, Doppelhofer and Miller (2004). This paper uses model weights built from information criteria and frequentist estimators of parameters to investigate the support for various determinants of growth in a cross-country study. Fernández, Ley and Steel (2001b), already discussed above with respect to computing the marginal likelihodd, can be viewed as a more formal inferential method for the same question. Another innovative application of variable selection with a *g-prior* is that to non-parametric regression in M. Smith and Kohn (1996).

A sensible approach to variable selection is to take advantage of MCMC output and particular shrinkage priors. A coarse name for one class of priors useful in variable selection is the spike-and-slab prior such as proposed in Mitchell and Beauchamp (1988). These priors place considerable weight or mass on the point of exclusion, where the regression coefficient equals zero (the *spike*), but are fairly uninformative about all other values (the *slab*). Although Mitchell and Beauchamp (1988) first proposed an uninformative prior for the slab, it is now common to have a slab component of the prior that is proper but relatively uninformative. An alternative proposed by Brown and Griffin (2010) has, as a special case, a generalization of the Bayesian Lasso of Park and Casella (2008).

George and McCulloch (1993) develop a computationally more efficient algorithm than the approach of Mitchell and Beauchamp (1988), called Stochastic Search Variable Selection (SSVS), using a mixture of normal priors. This approach does not exclude parameters exactly, rather it heavily shrinks unimportant parameters towards zero. Extending this to a high-dimensional setting, George, Sun and Ni (2008) adapt this prior and the SSVS algorithm for the problem of imposing a degree of parsimony upon a vector autoregressive model (see also Jochmann, Koop, León-González and Strachan (2013)). Koop, León-González and Strachan (2009) apply this approach to selecting which parameters vary – the mean equation coefficients, the structural parameters or the volatilities – in a time varying parameter VAR study of the evolution of monetary policy transmission.

Moving back from variable selection to direct computation of the marginal likelihood and the role of MCMC output, we discuss here a number of approaches to approximating the marginal likelihood that use the properties of the posterior to suggest estimators. With the advent of MCMC algorithms, it seemed natural to use the sequence of draws from the posterior to approximate the marginal likelihood. An early approach that used the MCMC output, but which has largely fallen from favour, is the Harmonic mean of the likelihood. This estimator, the average of the inverse of the likelihood, was presented in Newton and Raftery (1994) and, in the discussion following that article, Neal (1998) pointed out significant limitations of this approach. Most seriously, the estimator often has infinite variance (Chib, 1995) such that the number of draws from the posterior *" . . . required for this estimator to get close to the right answer will often be greater than the number of atoms in the observable universe."* (Neal, 2008). Modifications that improve upon this aspect have since been suggested.

Gelfand and Dey (1994) use the algebra of the integral to justify an alternative estimator that improved upon the Harmonic mean. They address the instability of the estimator by replacing the draws of the inverse of the likelihood by the ratio of a proper density with thin tails, over the product of the likelihood and the prior. A problem remained, however, in that it is sometimes difficult to judge *a priori* whether the chosen density will have tails sufficiently thin. A solution to this problem was proposed by Geweke (1999). Geweke's solution was to use a multivariate truncated normal as the weighting density, where truncation is to an ellipsoid with known volume. The tails were, by construction, therefore thin (in fact zero outside the ellipsoid). Also using MCMC output and very simple algebra of the identity relating the marginal likelihood to the construction of the posterior, Chib (1995) presents a simple estimator of the marginal likelihood. This estimator requires a value of the parameter vector in a high mass area (some value that is contained within a small Higher Posterior Density region) and an estimate of the full posterior. As the full posterior is unknown (hence our need to estimate the marginal likelihood), the MCMC output can be used in a Rao-Blackwellization of the appropriate conditionals to provide an estimate of this component. Perhaps because it does not require choosing a candidate density, and relies upon simple computations, Chib's method has been more widely adopted than the Gelfand-Dey approach or the Geweke improvement.

The approaches described above can be applied to estimate the marginal likelihood without reference to another model. However, we often wish to compare models and so are interested in the Bayes factor, and here we are interested in the case where these models nest within an encompassing model. A simple example of this is in variable selection, where the encompassing model has all available regressors of interest and the nested models have only a subset of regressors. In this fortunate situation, this nesting feature can be exploited to produce estimates of the Bayes factor and this can be achieved only using output from the encompassing model. An approach that does this is the Savage-Dickey-Density-Ratio proposed in Verdinelli and Wasserman (1995) (and see Dickey, 1971 for an early presentation of the conditions for this approach). To employ this tool, however, a condition must be met that the conditional prior for the unrestricted parameters conditional upon the restricted parameters at the point of the restriction, equals the prior for the unrestricted parameters in the restricted model. As convoluted as this condition sounds, it is generally met trivially. For example, independent priors will meet this condition. This technique, therefore, has a reasonably wide range of potential applications and we point to a few applications such as Koop and Poirier (2001), Fernández, León, Steel and Vázquez-Polo (2004), Koop, León-González and Strachan (2010), Kleibergen and Mavroeidis (2014) and Jarociński and Maćkowiak (2017).

9.6 Applications with New Methods

This chapter has, to this point, focused upon techniques without much consideration of the applications in which they have been employed. It is rarely the case, however, that econometricians develop solutions without first facing problems that need solving. As such, many methods are developed to address problems specific to models or applications. Or to address new questions in new (often more complex) model specifications for which satisfactory estimation techniques do not exist. Similarly, as has been a theme of this chapter, existing methods find uses in areas where they have not previously been employed. To fully explore the range of examples of this link would require a chapter on its own. Instead, this section points to applications in various areas of econometrics and the models or methods they used. In doing so, this will hopefully give some idea of the association between methods and models.

In macroeconometrics, Bayesian methods have found broad use in multivariate time series models. An important mulivariate model that has found considerable use in macroeconomics is the vector autoregression model (VAR) introduced into econometrics by Christopher Sims (Sims, 1980) and Bayesian methods were developed early for this model (Litterman 1980, 1986, Doan, Litterman & Sims, 1984, Kadiyala & Karlsson, 1997, and Sims & Zha, 1998).

Within the study of the transmission of monetary policy, Cogley and Sargent (2002, 2005) developed the time-varying parameter VAR to capture the changing economic and policy environment. These two models generated much discussion about the role of model specification upon the empirical results and subsequent conclusions about monetary policy. Many issues were addressed with the publication of Primiceri (2005), which remains a gold-standard for commencing inference with this model.

While they are very flexible and can capture an incredibly rich range of dynamics, VARs have long been criticised for not being parsimonious. The dimension of the parameter space for a VAR grows at rate of the number of variables squared, and so many studies employed models with relatively small number of variables (e.g., it was often the case that studies used no more than 10 variables in a system). Complex and over parameterized models have been eschewed by forecasters as they are generally regarded as likely to produce high levels of uncertainty around their forecasts. Bańbura, Giannone and Reichlin (2010) provided convincing evidence that this opinion was not necessarily well founded, provided the appropriate priors (shrinkage priors) were used. Their study compared VARs for 3, 7, 20 and 131 variables and found forecast performance (measured by mean square forecast error) improved with larger models. Importantly, they found that much of the gain in forecast improvement was achieved once the number of variables reached 20. As an aside, it is worth mentioning that this work provided an approach to mitigating another issue in VARs, that of possible nonfundamentalness (Hansen & Sargent, 1980).

Another important model in macroeconomics is the dynamic stochastic general equilibrium (DSGE) model. This macroeconomic model is built from microeconomic foundations and econometric inference on this model has benefited greatly by contributions from Bayesian methods. In an early attempt to improve and modernize policy advice at the Federal Reserve Bank, Del Negro and Schorfheide (2004) use a New Keynesian DSGE model as a prior for a Bayesian VAR. This work was followed by many new papers (e.g., Smets & Wouters, 2007, An & Schorfheide, 2007, and Adolfson, Laséen, Lindé & Villani, 2007) developing this area of application.

Bayesian methods have similarly proven useful in financial econometrics. An important component of financial processes is their volatility with its attendant relationship with risk. Two important papers that have improved modelling and inference on volatility include Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998). Both consider stochastic volatility and demonstrate new methods and improved inference. There have since been a great many applications and new methods in finance using Bayesian methods, to the point that an entire chapter of a volume (Johannes and Polson (2010)) is devoted specifically to the application of MCMC methods in continuous-time asset pricing models. For a wide ranging overview of developments in the literature on Bayesian methods in finance, we refer the reader to the recent special issue of the *Journal of Econometrics* (Kumbhakar & Mallick, 2026).

In developing their approach to computing the stochastic volatility model, Kim et al. (1998) use a mixture of normals to approximate the distribution for an error term. While the technique of using mixtures is not unique to Bayesian analysis, innovative methods have been developed for their computation that have expanded the range of applications. The stochastic volatility is one example of the use of mixtures in Bayesian analysis with another major area of application being in Bayesian microeconometrics. Finite Gaussian mixtures have found considerable application since early work by Diebolt and Robert (1994) with many extensions including the smoothly mixing regressions of Geweke and Keane (2007) and Villani, Kohn and Giordani (2009). Many mixtures take advantage of a Dirichlet prior distribution as discussed in Leslie, Kohn and Nott (2007) and P. J. Green and Richardson (2001). Excellent sources for learning about finite mixtures are Frühwirth-Schnatter (2006) and Fruhwirth-Schnatter, Celeux and Robert (2019).

There has been a long and rich relationship between microeconometrics and the Bayesian approach using tools such as data augmentation and with a general treatment of latent variables. Processes relevant to microeconomics have been studied by Bayesians since the 1980s. For example, an early econometric applications were to dichotomous qualitative response models in Zellner and Rossi (1984) and to the Tobit model in Chib (1992). Other contributions include Albert and Chib (1993), Dellaportas and Smith (1993). We refer to Chapter 6 of M. Li and Tobias (2011) for a useful overview. In addition to the limited dependent variables (including binary, multinomial), Bayesian methods have found application in an incredibly wide range of applications including duration models (Campolieti, 1997, and M. Li, 2006), hierarchical models (Lindley and Smith (1972), A. Smith (1973)), panel data (Chib, Greenberg and Winkelmann (1998), Chib and Jacobi (2007)), choice modelling, count and binary data (Burda, Harding & Hausman, 2008), ordinal data (Rossi, Allenby & McCulloch, 2012), treatment effects (Heckman, Lopes & Piatek, 2014, Chib & Jacobi, 2007), hurdle/sample selection models (van Hasselt, 2011), and many more.

Many of the same modelling approaches used in Bayesian microeconometrics, and some in macroeconometrics, can be found in the econometrics of marketing. A very early Bayesian contribution specifically to marketing is P. E. Green and Frank (1966). A number of useful reference sources include Rossi et al. (2012) and Rossi, Allenby and Misra (2024).

9.7 Conclusion

The Bayesian statistical method lay relatively dormant for the first almost 200 years after Bayes and Price (1763) and Laplace (1986). While it began to attract new attention in the early stages of the 20^{th} century, it was not until the later half of the century that econometrics began to fully embrace it. From the 1980s and 1990s, the number of Bayesian applications in econometrics grew exponentially as methods and computing power grew apace. This chapter has shown the process was not a straightforward one of methods being developed, employed for a while and then usurped by newer and better methods. Rather, methods often fell out of favour (such as numerical integration, the Laplace approximation and importance sampling) only to find new use later, often in combination with new methods (with, for example, particle filters and INLA).

Today, there remain few areas of econometrics where the Bayesian approach has not been applied. It is likely, however, that econometrics, and therefore Bayesian econometrics will find new uses and in many of these cases, future problems will likely be addressed by adapting existing methods to new problems. At the time of writing, we are seeing the rapid development of large language models in the implementation of AI. It is not clear at this point whether the effect of these models will be to increase our capacity to conduct research and so open new opportunities for research, or lead to new problems for economists, and therefore econometricians, to address. It is likely we will see both of these outcomes. Already, the Bayesian approach has begun to influence the functioning of large language models as evidenced by Qiu et al. (2026) who show how a Bayesian approach can improve their ability to update their beliefs.

In one of his last publications, Herman van Dijk considered the future of Bayesian econometrics. Specifically, in van Dijk (2024) he identified six challenges for society that will provide research opportunities. The importance of Bayesian econometrics, in its many forms, will continue to provide new answers to existing and previously unseen problems.

References

Adolfson, M., Laséen, S., Lindé, J. & Villani, M. (2007). Bayesian estimation of an open economy dsge model with incomplete pass-through. *Journal of International Economics*, *72*(2), 481-511.

Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669-679.

An, S. & Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews*, *26*(2-4), 113–172. Retrieved from https://doi.org/10.1080/07474930701220071 doi: 10.1080/07474930701220071

Ardia, D., Baştürk, N., Hoogerheide, L. & van Dijk, H. K. (2012). A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Computational Statistics & Data Analysis*, *56*(11), 3398-3414. (1st issue of the Annals of Computational and Financial Econometrics Sixth Special Issue on Computational Econometrics)

Arellano, M. & Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, *77*(2), 489–536. Retrieved 2026-06-15, from http://www.jstor.org/stable/40263873

Ausín, M. C. & Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, *51*(5), 2636-2652.

BaŞtürk, N., Çakmaklı, C., Ceyhan, S. P. & van Dijk, H. K. (2014). On the rise of Bayesian econometrics after Cowles Foundation Monographs 10, 14. *Œconomia*, *4*(3), 381-447.

Bauwens, L. (1984). *Bayesian full information analysis of simultaneous equation models using integration by Monte Carlo* (1st ed., Vol. 232). Berlin, Heidelberg: Springer-Verlag.

Bauwens, L. & Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. *The Econometrics Journal*, *1*, C23-C46.

Bayes, T. & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. (Reprinted in Biometrika 45 (1958), 293–315)

Bańbura, M., Giannone, D. & Reichlin, L. (2010). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Besag, J. & Green, P. J. (1993, 09). Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(1), 25-37.

Brown, P. J. & Griffin, J. E. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, *5*, 171-188.

Burda, M., Harding, M. & Hausman, J. (2008). A Bayesian mixed logit-probit model for multinomial choice: Estimating demand systems and measuring consumer preferences. *Journal of econometrics*, *147*(2), 232-246.

Byrne, S. & Giralomi, M. (2013). Geodesic Monte Carlo on Embedded Manifolds. *Scandinavian Journal of Statistics*, *40*(4), 825–845.

Campolieti, M. (1997). Bayesian estimation of duration models: An application of the multiperiod probit model . *Empirical Economics*, *22*, 461-480.

Chan, J. C. C., Glynn, P. W. & Kroese, D. P. (2011). A comparison of cross-entropy and variance minimization strategies. *Journal of Applied Probability*, *48*(A), 183–194. doi: 10.1239/jap/1318940464

Chan, J. C. C. & Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*, 101-120.

Chan, J. C. C., Koop, G., Poirier, D. J. & Tobias, J. L. (2020). *Bayesian econometric methods*. Japan: Produced by Amazon.

Chan, J. C. C., Koop, G. & Potter, S. M. (2013). A New Model of Trend Inflation. *Journal of Business & Economic Statistics*, *31*(1), 94–106.

Chan, J. C. C., Koop, G. & Potter, S. M. (2016). A Bounded Model of Time Variation in Trend Inflation, Nairu and the Phillips Curve. *Journal of Applied Econometrics*, *31*(3), 551-565.

Chan, J. C. C. & Kroese, D. P. (2012). Improved cross-entropy method for estimation. *Statistics and computing*, *22*(5), 1031-1040.

Chan, J. C. C., León-González, R. & Strachan, R. W. (2018). Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association*, *113*, 819-828.

Chan, J. C. C. & Strachan, R. W. (2012). *Estimation in non-linear non-Gaussian state-space models with precision-based methods* (Working Paper No. Working Paper No. 2012-13). CAMA Working Paper Series.

Chan, J. C. C. & Strachan, R. W. (2014). *The zero lower bound: Implications for modelling the interest rate* (Working Paper No. Working Paper No. 42_14). Rimini Centre for Economic Analysis.

Chetty, V. K. (1968a). Bayesian Analysis of Haavelmo's Models. *Econometrica*, *36*(3/4), 582–602.

Chetty, V. K. (1968b). Pooling of time series and cross section data. *Econometrica*, *36*(2), 279–290. Retrieved 2026-05-28, from http://www.jstor.org/stable/

1907490

Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics*, *51*(1), 79-99.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, *90*, 1313-1321.

Chib, S. (2011). Introduction to Simulation and MCMC Methods. In J. Geweke, G. Koop & H. Van Dijk (Eds.), *The Oxford Handbook of Bayesian Econometrics.* Oxford University Press.

Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, *49*(4), 327–335.

Chib, S., Greenberg, E. & Winkelmann, R. (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics*, *86*(1), 33–54.

Chib, S. & Jacobi, L. (2007). Modeling and calculating the effect of treatment at baseline from panel outcomes. *Journal of Econometrics*, *140*(2), 781–801.

Choi, J., Jo, S. & Kim, J. (2025). A Bayesian Time-Varying Coefficient Model for Cobb–Douglas Production Function. *Computational Economics*, *65*, 1429–1455.

Cogley, T. & Sargent, T. J. (2002, October). *Evolving Post-World War II US Inflation Dynamics* (None ed., Vol. None). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/h/nbr/nberch/11068.html

Cogley, T. & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262-302. (Monetary Policy and Learning)

Congdon, P. (2007). *Bayesian statistical modelling*. Wiley. Retrieved from https://books.google.com.au/books?id=PthTOJyGGNIC

Cowles, M. K. & Carlin, B. P. (1996, 06). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, *91*(434), 883-904.

Cowles, M. K. & Rosenthal, J. S. (1998). A simulation approach to convergence rates for Markov chain Monte Carlo algorithms. *Statistics and Computing*, *8*(2), 115–124.

Creal, D. (2012). A Survey of Sequential Monte Carlo Methods for Economics and Finance. *Econometric Reviews*, *31*(3), 245–296.

de Finetti, B. (1937). Foresight: Its logical laws, its subjective sources. *Annales de l'Institut Henri Poincaré*, *7*, 1–68. (English translation published in 1964 in *Studies in Subjective Probability*, edited by H.E. Kyburg and H.E. Smokler)

Dellaportas, P. & Smith, A. (1993). Bayesian inference for generalized linear and proportional hazards models via gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *42*(3), 443–459.

Del Negro, M. & Schorfheide, F. (2004). Priors from General Equilibrium Models for VARS. *International Economic Review*, *45*(2), 643-673.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Dickey, J. M. (1971). The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics*, *42*(1), 204–223.

Dickey, J. M. & Lientz, B. P. (1970). The Weighted Likelihood Ratio, Sharp Hypotheses about Chances, the Order of a Markov Chain. *The Annals of Mathematical Statistics*, *41*(1), 214–226.

Diebolt, J. & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(2), 363–375.

Doan, T., Litterman, R. & Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, *3*, 1-100.

Doucet, A. (2026). *Sequential Monte Carlo Methods & Particle Filters Resources.* https://www.stats.ox.ac.uk/~doucet/smc_resources.html. (Accessed: 2026-06-27)

Droumaguet, M., Warne, A. & Woźniak, T. (2017). Granger Causality and regime inference in Markov switching VAR models with Bayesian methods. *Journal of Applied Econometrics*, *32*(4), pp. 802–818.

Drèze, J. H. (1976). Bayesian Limited Information Analysis of the Simultaneous Equations Model. *Econometrica*, *44*(5), 1045–1075.

Drèze, J. H. (1977). Bayesian regression analysis using poly-t densities. *Journal of Econometrics*, *6*(3), 329-354.

Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216-222.

Durbin, J. & Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Eisenstat, E. & Strachan, R. W. (2026). Singular vector autoregressions. *Journal of Econometrics*, *256*, 106255. Retrieved from https://www.sciencedirect.com/science/article/pii/S030440762600076X doi: https://doi.org/10.1016/j.jeconom.2026.106255

Fan, Y., Brooks, S. P. & Gelman, A. (2006). Output Assessment for Monte Carlo Simulations via the Score Statistic. *Journal of Computational and Graphical Statistics*, *15*(1), 178–206.

Fernández, C., Ley, E. & Steel, M. F. (2001a). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*(2), 381-427.

Fernández, C., Ley, E. & Steel, M. F. J. (2001b). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, *16*(5), 563–576.

Fernández, C., León, C. J., Steel, M. F. J. & Vázquez-Polo, F. J. (2004). Bayesian analysis of interval data contingent valuation models and pricing policies. *Journal of Business & Economic Statistics*, *22*(4), 431–442.

Fienberg, S. (2006, 03). When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, *1*. doi: 10.1214/06-BA101

Friedman, M. (1953). The methodology of positive economics. *Essays in positive economics*, *3*(3), 145–178.

Frühwirth-Schnatter, S. (1999). *Model likelihoods and Bayes factors for switching and mixture models* (Technical Report Nos. 1999-71, revised version January 2001). Vienna, Austria: Vienna University of Economics and Business Administration.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.

Frühwirth-Schnatter, S. (2019). Keeping the balance: Bridge sampling for marginal likelihood estimation in finite mixture, mixture of experts and Markov mixture models. *Barzilian Journal of Probability and Statistics*, *33*, 706-733.

Fruhwirth-Schnatter, S., Celeux, G. & Robert, C. E. (Eds.). (2019). *Handbook of Mixture Analysis (1st ed.).* Chapman and Hall/CRC.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.

Gelfand, A. & Dey, D. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, 501-514.

Gelfand, A. & Smith, A. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, *85*(410), 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). *Bayesian data analysis*. Boca Raton: CRC Press.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Geman, S. & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721-741. doi: 10.1109/TPAMI.1984.4767596

George, E. I. & McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, *88*(423), 881–889.

George, E. I., Sun, D. & Ni, S. (2008). Stochastic Search Model Selection for Restricted VAR Models. *Journal of Econometrics*, *142*, 553–580.

Geweke, J. (1988). The Secular and Cyclical Behavior of Real GDP in 19 OECD Countries, 1957-1983. *Journal of Business & Economic Statistics*, *6*(4), 479–486.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, *57*(6), 1317-1339.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid & A. Smith (Eds.), *Bayesian statistics 4* (p. 169-193). Oxford: Oxford University Press.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: inference, development, and communication. *Econometric Reviews*, *18*(1), 1–73.

Geweke, J. (2001). Bayesian econometrics and forecasting. *Journal of Econometrics*, *100*(1), 11-15. (Open Forum on the Current State and Future Challenges of Economics)

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics.* Hoboken, N.J: John Wiley.

Geweke, J. & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130-141. (Annals Issue on Forecasting)

Geweke, J. & Keane, M. (2007). Smoothly mixing regressions. *Journal of Econometrics*, *138*(1), 252–290.

Geweke, J., Koop, G. & Van Dijk, H. (2011). *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.

Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In S. Brooks, A. Gelman, G. Jones & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 3–48). New York: Chapman and Hall/CRC.

Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables. (1999). *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *61*(2), 331–344.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov chain monte carlo in practice*. London: Chapman & Hall/CRC.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *41*(2), 337–348.

Gordon, N., Salmond, J. & Smith, A. (1993). A novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process*, *140*, 107-113.

Green, P. E. & Frank, R. E. (1966). Bayesian Statistics and Marketing Research. *Applied Statistics*, *15*(3), 173-190.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, *82*(4), 711–732.

Green, P. J. & Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian journal of statistics*, *28*(2), 355–375.

Griffin, J. E. & Steel, M. F. J. (2008). Flexible mixture modelling of stochastic frontiers. *Journal of Productivity Analysis*, *29*(1), 33–50.

Gûnel, E. & Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, *61*(3), 545–557. Retrieved 2026-04-16, from http://www.jstor.org/stable/2334738

Hammersley, J. M. & Handscomb, D. C. (1964). *Monte Carlo Methods.* New York: Wiley.

Han, J. & Lee, Y. (2024). Enhanced Laplace approximation. *Journal of Multivariate Analysis*, *202*, 105321.

Handschin, J. E. & Mayne, D. Q. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering† . *International Journal of Control*, *9*(5), 547–559.

Hansen, L. P. & Sargent, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control*, *2*, 7-46.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*(1), 97–109.

Heckman, J. J., Lopes, H. F. & Piatek, R. (2014). Treatment Effects: A Bayesian Perspective. *Econometric Reviews*, *33*(1-4), 36–67.

Herce, A. & Salvador, M. (2026). Bayesian gibbs slice sampler: a novel approach to efficient mcmc. *Computational economics*.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business* & *Economic Statistics*, *12*, 371-418.

Jarociński, M. & Maćkowiak, B. (2017). Granger causal priority and choice of variables in vector autoregressions. *The Review of Economics and Statistics*, *99*(2), pp. 319–329.

Jeffreys, H. (1939). *Theory of probability*. Oxford: Clarendon Press.

Jochmann, M., Koop, G., León-González, R. & Strachan, R. (2013). Stochastic search variable selection in vector error correction models with an application to a model of the UK macroeconomy. *Journal of Applied Econometrics*, *28*(1), 62–81.

Johannes, M. & Polson, N. (2010). CHAPTER 13 - MCMC Methods for Continuous-Time Financial Econometrics. In Y. Aït-Sahalia & L. P. Hansen (Eds.), *Handbook of Financial Econometrics: Applications* (Vol. 2, p. 1-72). San Diego: Elsevier.

Kadiyala, K. R. & Karlsson, S. (1997). Numerical mehtods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, *12*, 99-132.

Kalli, M., Walker, S. G. & Damien, P. (2013). Modeling the Conditional Distribution of Daily Stock Index Returns: An Alternative Bayesian Semiparametric Model. *Journal of Business & Economic Statistics*, *31*(4), 371–383.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, *65*, 361-393.

Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, *5*(1), 1–25.

Kleibergen, F. & Mavroeidis, S. (2014). Identification issues in limited-information Bayesian analysis of structural macroeconomic models. *Journal of Applied Econometrics*, *29*(7), 1183–1209.

Kloek, T. & van Dijk, H. K. (1978, January). Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. *Econometrica*, *46*(1), 1-19.

Koop, G. & Korobilis, D. (2010, 06). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, *3*(4), 267-358.

Koop, G., León-González, R. & Strachan, R. W. (2009, April). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, *33*(4), 997-1017.

Koop, G., León-González, R. & Strachan, R. W. (2010). Dynamic probabilities of restrictions in state space models: An application to the phillips curve. *Journal of Business & Economic Statistics*, *28*(3), 370–379.

Koop, G. & Poirier, D. J. (2001). Testing for optimality in job search models. *The Econometrics Journal*, *4*(2), 257–272.

Koop, G., Poirier, D. J. & Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge University Press.

Kreuzer, A. & Czado, C. (2021). Bayesian inference for a single factor copula stochastic volatility model using Hamiltonian Monte Carlo. *Econometrics and Statistics*, *19*, 130-150.

Kroese, D. P., Taimre, T. & Botev, Z. I. (2011). *Handbook for Monte Carlo methods* (1st ed., Vol. 706). Hoboken, N.J: John Wiley & Sons, Inc.

Kullback, S. (1959). *Information theory and statistics.* New York: Wiley.

Kumbhakar, S. C. & Mallick, S. K. (2026). Bayesian methods in economics and finance: A unified survey and taxonomy. *Journal of Econometrics*, 106269.

Lamnisos, D., Griffin, J. E. & Steel, M. F. J. (2009). Transdimensional Sampling Algorithms for Bayesian Variable Selection in Classification Problems With Many More Variables Than Observations. *Journal of Computational and Graphical Statistics*, *18*(3), 592–612.

Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical Science*, *1*(3), 364–378. Retrieved 2026-03-15, from http://www.jstor.org/stable/2245476

Lawrence, E., Liu, C., Bingham, D. & Nair, V. N. (2008). Bayesian Inference for Multivariate Ordinal Data Using Parameter Expansion. *Technometrics*, *50*(2), 182–191.

Leslie, D. S., Kohn, R. & Nott, D. J. (2007). A general approach to heteroscedastic linear regression. *Statistics and Computing*, *17*(2), 131–146.

Li, M. (2006). High school completion and future youth unemployment: New evidence from high school and beyond. *Journal of Applied Econometrics*, *21*(1), 23–53.

Li, M. & Tobias, J. (2011). Bayesian methods in microeconometrics. In J. Geweke, G. Koop & H. Van Dijk (Eds.), *The Oxford Handbook of Bayesian Econometrics.* Oxford University Press.

Li, Y., Mallick, S. K., Wang, N., Yu, J. & Zeng, T. (2025). Deviance Information Criterion for Bayesian model selection: Theoretical justification and applications. *Journal of Econometrics*, 105978.

Lindley, D. V. & Smith, A. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(1), 1–41.

Litterman, R. B. (1980). *A Bayesian procedure for forecasting with vector autoregressions.* Massachusetts Institute of Technology. (mimeo)

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business* & *Economic Statistics*, *4*, 25–38.

Liu, J. (1994). The Collapsed Gibbs Sampler in Bayesian Computations With Applications to Gene Regulation Problem. *Journal of the American Statistical Association*, *89*, 958-966.

Liu, J., Wong, W. & Kong, A. (1994, 03). Covariance Structure of the Gibbs Sampler with Applications to the Comparison of Estimators and Augmentation Schemes. *Biometrika*, *81*, 27-40.

Liu, J. & Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, *448*, 1264-1274.

Llorente, F., Martino, L., Delgado, D. & López-Santiago, J. (2023). Marginal likelihood computation for model selection and hypothesis testing: An extensive review. *SIAM Rev.*, *65*(1), 3–58.

Loddo, A., Ni, S. & Sun, D. (2011). Selection of Multivariate Stochastic Volatility Models via Bayesian Stochastic Search. *Journal of Business & Economic Statistics*, *29*(3), 342–355.

MacEachern, S. N. (2007). Comment on article by Jain and Neal. *Bayesian Analysis*, *2*(3), 483-494.

Madigan, D., York, J. & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review / Revue Internationale de Statistique*, *63*(2), 215–232.

Martin, G. M., Frazier, D. T. & Robert, C. P. (2024a). Approximating Bayes in the 21st Century. *Statistical Science*, *39*(1), 20 – 45.

Martin, G. M., Frazier, D. T. & Robert, C. P. (2024b). Computing Bayes: From Then 'Til Now. *Statistical Science*, *39*(1), 3-19. doi: https://doi.org/10.1214/22-STS876

McLure, M., Turkington, D. & Weber, E. J. (2010). A Conversation with Arnold Zellner. *History of Economics Review*, *51*(1), 72–81.

Meng, X.-L. & van Dyk, D. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, *86*, 301-320.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953, June). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, *21*(6), 1087-1092.

Mitchell, T. J. & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032.

Neal, R. M. (1998). *Erroneous results in 'Marginal likelihood from the Gibbs output'.* available at http://www.cs.toronto.edu/~radford/ftp/chib-letter.pdf.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, *31*(3), 705–741.

Neal, R. M. (2008). *Radford Neal's blog.* Retrieved 2008-08-17 at 12:09 am, from https://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever/

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 139–188). New York: Chapman and Hall/CRC.

Newton, M. A. & Raftery, A. E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(1), 3-26.

Ni, S. & Sun, D. (2003). Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models. *Journal of econometrics*, *115*(1), 159-197.

Nishimura, A., Dunson, D. B. & Lu, J. (2020). Discontinuous Hamiltonian Monte Carlo for discrete parameters and discontinuous likelihoods. *Biometrika*, *107*(2), 365–380.

Ogden, H. (2021). On the error in laplace approximations of high-dimensional integrals. *Stat*, *10*(1), e380.

Park, T. & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., ... Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705-871.

Phillips, P. C. (1995). Bayesian model selection and prediction with empirical applications. *Journal of Econometrics*, *69*(1), 289-331. (Annals of Econometrics: Bayesian and Classical Econometric Modelling of Time Series)

Phillips, P. C. & Ploberger, W. (1994, August). Posterior odds testing for a unit root with data-based model selection. *Econometric Theory*, *10*(3-4), 774-808.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, *72*(3), 821–852.

Qin, D. (1996). Bayesian econometrics: The first twenty years. *Econometric Theory*, *12*(3), 500–516.

Qiu, L., Sha, F., Allen, K., Yoon, K., Linzen, T. & van Steenkiste, S. (2026). Bayesian teaching enables probabilistic reasoning in large language models. *Nature Communications*, *17*, Article number 1238.

Raftery, A. E., Madigan, D. & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.

Raiffa, H. & Schlaifer, R. (1961). *Applied statistical decision theory.* Oxford: Harvard University Press.

Ramsey, F. P. (1964). Truth and probability. In H. E. Kyburg & H. E. Smokler (Eds.), *Studies in subjective probability* (pp. 61–92). New York: John Wiley & Sons. (Reprinted from 1926)

Rebeschini, P. & van Handel, R. (2015, October). Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, *25*(5).

Ritter, C. & Tanner, M. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, *87*, 861-868.

Robert, C. P. & Casella, G. (2005). *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Berlin, Heidelberg: Springer-Verlag.

Robert, C. P. & Casella, G. (2011). A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science*, *26*(1), 102–115.

Rossi, P. E., Allenby, G. M. & McCulloch, R. (2012). *Bayesian, statistics and marketing* (1st ed. ed.). Chichester: John Wiley and Son.

Rossi, P. E., Allenby, G. M. & Misra, S. (2024). *Bayesian statistics and marketing* (Second edition ed.). Hoboken, NJ: Wiley.

Roy, V. (2020). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, *7*(1), 387–412.

Rue, H., Martino, S. & Chopin, N. (2009, 04). Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations.

Journal of the Royal Statistical Society Series B: Statistical Methodology, *71*(2), 319-392.

Sala-i-Martin, X., Doppelhofer, G. & Miller, R. I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *The American Economic Review*, *94*(4), 813–835.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Shun, Z. & McCullagh, P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(4), 749–760.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*, 1-48.

Sims, C. A. (2001). Comment on Cogley and Sargent's' Evolving post World War II U.S. inflation dynamics. In *Nber macroeconomics annual* (Vol. 16, pp. 373–379). MIT Press.

Sims, C. A. & Zha, T. (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, *39*(4), 949–968.

Smets, F. & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *The American Economic Review*, *97*, 586-606.

Smith, A. (1973). A general bayesian linear model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *35*(1), 67–75.

Smith, A. & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, *55*, 3-23.

Smith, M. & Kohn, R. (1996). Nonparametric regression using bayesian variable selection. *Journal of Econometrics*, *75*(2), 317-343.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, *64*(4), 583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *76*(3), 485–493.

Steel, M. F. J. (2011, 01). 8. bayesian model averaging and forecasting. *Bulletin of EU and US Inflation and Macroeconomic Analysis*, *200*.

Steel, M. F. J. (2020). Model averaging and its use in economics. *Journal of Economic Literature*, *58*(3), pp. 644–719.

Stein, C. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Proceedings of the third berkeley symposium on mathematical statistics and probability* (Vol. I, p. 197-206). Berkeley: Statistical Laboratory of the University of California, Berkeley.

Stigler, S. M. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*(3), 359–363.

Strachan, R. W. & Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, *123*, 307-325.

Tanner, M. A. (1993). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (2nd ed.). New York:

Springer-Verlag.

Tanner, M. A. & Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, *22*, 1701-1762.

Tierney, L. & Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, *81*(393), 82–86.

Tierney, L., Kass, R. E. & Kadane, J. B. (1989). Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions. *Journal of the American Statistical Association*, *84*(407), 710–716.

van Hasselt, M. (2011). Bayesian inference in a sample selection model. *Journal of Econometrics*, *165*(2), 221-232.

van Dijk, H. K. (2024). Challenges and opportunities for twenty first century bayesian econometricians: A personal view. *Studies in Nonlinear Dynamics & Econometrics*, *28*(2), 155–176.

van Dyk, D. A. & Park, T. (2008). Partially Collapsed Gibbs Samplers. *Journal of the American Statistical Association*, *103*(482), 790–796.

Verdinelli, I. & Wasserman, L. (1995). Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, *90*(430), 614–618.

Villani, M., Kohn, R. & Giordani, P. (2009). Regression density estimation using smooth adaptive gaussian mixtures. *Journal of Econometrics*, *153*(2), 155-173.

West, M. & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (Second Edition. ed.). New York, NY: Springer.

Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, *146*(2), 329-341. (Honoring the research contributions of Charles R. Nelson)

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: J. Wiley.

Zellner, A. (2008, October). *Bayesian econometrics: past, present, and future.* Emerald Group Publishing Limited.

Zellner, A., Hill, e., R. Carter & Fomby, T. B. (1997). The Bayesian Method of Moments (BMOM): Theory and Applications. In *Applying maximum entropy to econometric problems. 1997, pp. 85-105* (p. 85-105).

Zellner, A. & Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, *25*(3), 365-393.

Zellner, A., Tobias, J. & Ryu, H. (1998). Bayesian Method of Moments (BMOM) Analysis of Parametric and Semiparametric Regression Models. *Research in Agricultural & Applied Economics*.