Bayesian Comparison of Econometric Models

John Geweke
University of Minnesota
Federal Reserve Bank of Minneapolis
geweke@atlas.socsci.umn.edu

First version: June 2, 1994
This revision: September 30, 1995

Abstract

This paper integrates and extends some recent computational advances in Bayesian inference with the objective of more fully realizing the Bayesian promise of coherent inference and model comparison in economics. It combines Markov chain Monte Carlo and independence Monte Carlo with importance sampling to provide an efficient and generic method for updating posterior distributions. It exploits the multiplicative decomposition of marginalized likelihood into predictive factors, to compute posterior odds ratios efficiently and with minimal further investment in software. It argues for the use of predictive odds ratios in model comparison in economics. Finally, it suggests procedures for public reporting that will enable remote clients to conveniently modify priors, form posterior expectations of their own functions of interest, and update the posterior distribution with new observations. A series of examples explores the practicality and efficiency of these methods.

This paper was initially prepared as the inaugural Colin Clark Lecture, Australasian Meetings of the Econometric Society, July 1994. I wish to acknowledge helpful comments made at this meeting, in seminars at Cambridge University, Federal Reserve System Board of Governors, Harvard-M.I.T., University of Kansas, University of Minnesota, Northwestern University, University of Pennsylvania, Princeton University, and University of Virginia, and at the 1994 summer and 1995 winter North American meetings of the Econometric Society, the 1994 North American and world meetings of the International Society for Bayesian Analysis, and the 1995 Bath international workshop on model comparison. The paper has benefited from discussions with Jim Berger, Gary Chamberlain, Jon Faust, Bill Griffiths, Peter Phillips, Christopher Sims, Mark Steel and Arnold Zellner. Remaining errors and shortcomings are entirely the author's. Zhenyu Wang provided research assistance. This work was supported in part by National Science Foundation Grant SES-9210070. The views expressed in this paper are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

Recent substantial advances in computational methods have greatly expanded our ability to apply Bayesian procedures in econometrics and other statistical sciences. Whereas just a few years ago applied Bayesian inference was limited to a few textbook models, in an increasing number of instances computationally intensive Bayesian methods are proving more practical and reliable than non-Bayesian procedures even by conventional non-Bayesian criteria. (Jacquier, Polson and Rossi (1994) provide such a comparison for stochastic volatility models, as do Geweke, Keane and Runkle (1994) for multinomial probit models.) These recent advances have exploited dramatic decreases in computation costs, and they are likely to continue as these costs decline even further.

This paper integrates and extends several of these advances, with the objective of realizing the promise of a complete and coherent framework for statistical inference that is inherent in Bayesian theory. It shows that some of the elegant constructions in Bayesian analysis are by no means limited to the elucidation of statistical principles. They also form the basis for the more mundane but essential task of efficient computation, and place the workaday business of diagnostics, outlier analysis, and model comparison on a sound yet practical footing. The paper breaks fresh ground in four directions.

First, the work introduces a combination of Markov chain Monte Carlo and independence Monte Carlo with importance sampling, including systematic procedures for the assessment of approximation error (Section 3.4). The extension is quite straightforward, yet it provides a widely applicable computational tool for the task of rapid updating of posterior distributions that has heretofore been unavailable.

Second, this research recapitulates the decomposition of marginalized likelihood as the product of predictive factors (Section 2.1). This decomposition forms the basis for the efficient computation of marginalized likelihoods -- and therefore Bayes factors and posterior odds ratios -- that has proven elusive and intractable (Sections 4.1 and 4.2). The predictive factors turn out to be precisely the importance sampling weights that are required in the combination of Markov chain Monte Carlo, and independence Monte Carlo with importance sampling, for the purposes of updating. Predictive likelihoods for individual observations provide diagnostics of model inadequacy and their ratios provide a useful analysis of posterior odds ratios (Section 5).

Third, this paper argues that in the construction of dynamic econometric models -- and, probably, many other kinds of statistical models as well -- there is often an identifiable portion of the sample that ought to be regarded as part of the prior (Section 2.2). This

argument provides a practical resolution of well established and widely known logical difficulties with respect to improper priors, "data mining," and public reporting. It is at least as logically appealing as alternative approaches (but perhaps no more likely to command a consensus).

Finally, the paper suggests specific standards for public reporting that exploit very recent and drastic declines in the costs of storing and communicating massive quantitative information. It argues (Section 4.3) that these standards will enable the prototypical remote client (Hildreth, 1963) to impose his or her subjective priors, investigate prior robustness, evaluate new loss or other functions of interest, and/or update the reported posterior with new observations -- all very rapidly with 1994 technology, and with simple generic software.

The technical details of these ideas require a different order of development. The next section reviews some classical compositions and decompositions of posterior odds ratios, and argues for a prior distribution based on an identifiable portion of the sample. Section 3 reviews recently developed simulation methods for the computation of posterior moments and introduces the modest extension just described. Section 4 takes up the important task of efficient computation of the elements of posterior odds ratios described in Section 2, and describes a public reporting format based on these methods. Examples in Sections 5 and 6 provide encouraging evidence on the practicality of the procedures proposed in the paper.

This paper argues for an agenda as much as it presents new results. The reader will note that some innovations in computation -- *e.g.*, updating by importance sampling rather than recomputation -- are not featured in the examples. (Forthcoming revisions will take up nontrivial examples of these procedures.) More generally, the methods proposed here ultimately require repeated application to actual problems for a complete assessment of their utility.

2. Posterior odds and Bayesian practice

The posterior odds ratio is a well established concept for model comparison. It constitutes the fundamental means of model comparison in subjective Bayesian analysis, and is central to the classical expected utility theory of decision making under uncertainty. (DeGroot (1970) provides a detailed fundamental argument.) Here we recapitulate this principle to establish notation, and introduce some compositions and decompositions of the posterior odds ratio that form the basis for the rest of this paper. We argue that one of these decompositions often provides a good formal model of the process of model construction in economics.

2.1 Priors, marginalized likelihoods, Bayes factors, and posterior odds ratios

Let $\{y_t\}_{t=1}^{T}$ be a set of observations whose conditional densities $y_t|(y_1,\mathbb{K},y_{t-1},\theta)$ under model j are given by $\mathrm{f}_{jt}(y_t|Y_{t-1},\theta_j)$, where $Y_t \equiv \{y_s\}_{s=1}^{t}, Y_0 = \{\varnothing\}$ and θ_j is the vector of parameters in model j. The prior probability of model j is p_j and conditional on model j the prior density kernel for θ_j is $\mathrm{f}_{j0}(\theta_j)$. If the prior density is proper then the kernel is taken to be the density itself. Let

$$\mathrm{L}_{jt}(\theta_j, Y_t) = \prod_{s=1}^{t} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)$$

denote the partial likelihood through observation t. Then conditional on model j and Y_t the posterior density for θ_j is

$$\mathrm{p}_{jt}(\theta_j|Y_t) = \mathrm{f}_{j0}(\theta_j)\mathrm{L}_{jt}(\theta_j,Y_t) \Big/ \int_{\Theta_j} \mathrm{f}_{j0}(\theta_j)\mathrm{L}_{jt}(\theta_j,Y_t)d\theta_j \qquad (1)$$

so long as the integral in the denominator converges. The marginalized likelihood for model j and the subsample for observations 1 through t is

$$M_{jt} = \int_{\Theta_j} \mathrm{f}_{j0}(\theta_j)\mathrm{L}_{jt}(\theta_j,Y_t)d\theta_j \qquad (2)$$

provided the prior is proper. The Bayes factor in favor of model j versus model k, given observations 1 through t, is $B_{j|k,t} = M_{jt}/M_{kt}$, and the posterior odds ratio in favor of model j versus model k, given observations 1 through t, is

$$POR_{j|k,t} = p_j M_{jt}/p_k M_{kt} = (p_j/p_k)B_{j|k,t}.$$

The concepts of predictive likelihood, predictive Bayes factor, and predictive odds ratio are closely related. The predictive likelihood for observations $u+1$ through t, given model j and observations 1 through u, is

$$\hat{p}_{ju}^{t} \equiv \int_{\Theta_j} \mathrm{p}_{ju}(\theta_j|Y_u)\prod_{s=u+1}^{t} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)d\theta_j . \qquad (3)$$

The predictive Bayes factor is $\hat{B}_{j|k,u}^{t} = \hat{p}_{ju}^{t}/\hat{p}_{ku}^{t}$, and the predictive odds ratio is $PRED_{j|k,u}^{t} = p_j\hat{p}_{ju}^{t}/p_k\hat{p}_{ku}^{t}$, both in favor of model j versus model k for observations $u+1$ through t. These decompositions are well known.

Using (1) to substitute for $\mathrm{p}_{ju}(\theta_j|Y_u)$ in (3),

$$\hat{p}_{ju}^{t} = \int_{\Theta_j} \left\{ \frac{\mathrm{f}_{j0}(\theta_j)\prod_{s=1}^{u} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)}{\int_{\Theta_j} \mathrm{f}_{j0}(\theta_j)\prod_{s=1}^{u} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)d\theta_j} \right\} \prod_{s=u+1}^{t} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)d\theta_j$$

$$= \frac{\int_{\Theta_j} \mathrm{f}_{j0}(\theta_j)\prod_{s=1}^{t} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)d\theta_j}{\int_{\Theta_j} \mathrm{f}_{j0}(\theta_j)\prod_{s=1}^{u} \mathrm{f}_{js}(y_s|Y_{s-1},\theta_j)d\theta_j} = \frac{M_{jt}}{M_{ju}}.$$

Hence for any $0 \le u = s_0 < s_1 < \mathbb{K} < s_q = t$, we have

$$\hat{p}_{ju}^{t} = \frac{M_{js_1}}{M_{js_0}} \cdot \frac{M_{js_2}}{M_{js_1}} \cdot \mathbb{L} \cdot \frac{M_{js_q}}{M_{js_{q-1}}} = \prod_{\tau=1}^{q} \hat{p}_{js_{\tau-1}}^{s_\tau} \qquad (4)$$

The predictive likelihood is thereby expressed as a product of *linked predictive likelihoods*. The decomposition (4) may be interpreted as a succession of q multiplicative updatings to the marginalized likelihood $\hat{p}_{j_0}^{s_0}$ at time s_0 that lead to the marginalized likelihood $\hat{p}_{j_0}^{s_q}$ at time s_q. The particular case $s_0 = 0, s_q = t$ provides a decomposition of the marginalized likelihood for the entire sample, and if $s_\mathbb{1} - s_{\mathbb{1}-1} = 1 \; \forall \; \mathbb{1}$ the decomposition is complete. The decomposition (4) is of interest as a model diagnostic, especially the complete decomposition: an unusually low value of $\hat{p}_{s_{\mathbb{1}-1}}^{s_\mathbb{1}} = \hat{p}_{s_\mathbb{1}-1}^{s_\mathbb{1}}$ indicates that observation $s_\mathbb{1}$ is improbable conditional on model j and the previous observations.

There are corresponding decompositions of the predictive Bayes factor,

$$\hat{B}_{j|k,s_0}^{s_q} = \hat{p}_{js_0}^{s_q} \Big/ \hat{p}_{ks_0}^{s_q} = \prod_{\mathbb{1}=1}^{q} \left(\hat{p}_{js_{\mathbb{1}-1}}^{s_\mathbb{1}} \Big/ \hat{p}_{ks_{\mathbb{1}-1}}^{s_\mathbb{1}} \right) = \prod_{\mathbb{1}=1}^{q} \hat{B}_{j|k,s_{\mathbb{1}-1}}^{s_\mathbb{1}} .$$

This decomposition can indicate observations, or groups of observations, that are more probable under one model or the other. It can lead to the identification of observations that are decisive in Bayes factors that are quite large or small.

2.2 Variants and alternatives

For a given data set, likelihood function and prior distribution, the posterior distribution (1) and the marginalized likelihood (2) are sufficient. In particular, these objects are all that is required to compare the model at hand with other models, including models not yet conceived and those which neither nest nor are nested in the model at hand. Posterior odds ratios establish the posterior probabilities of models conditional on a set of models, and together with the posterior distribution for each model this information is sufficient for formal decision making.

Despite these advantages marginalized likelihoods are not widely used as sufficient summary statistics for Bayesian model comparison. A key difficulty is that the marginalized likelihood is defined by the prior distribution as well as the likelihood function, and there is rarely a single specific proper prior distribution on which most investigators would agree. For the purposes of expressing a posterior distribution this problem is frequently obviated by the use of uninformative prior distributions. In particular, Jeffreys (1961) suggested improper prior distributions to represent knowing little, that are frequently employed when reporting posterior moments (Zellner, 1971, pp. 40-53). This does not resolve the difficulty with respect to posterior odds ratios, however, because proper

prior distributions are essential not only to the interpretation of the posterior odds ratio but also to the construction of the marginalized likelihood. For suppose that $\left\{\mathrm{f}_{j0}^{(n)}\left(\theta_j\right)\right\}$ is a sequence of proper prior densities with the property $\lim_{n\to\infty}\mathrm{f}_{j0}^{(n)}\left(\theta_j\right)=0\ \forall\ \theta_j\in\Theta_j$. Let $\left\{\mathrm{p}_{jt}^{(n)}\left(\theta_j|Y_t\right)\right\}$ denote the corresponding sequence of posterior densities, and $\left\{M_{jt}^{(n)}\right\}$ the corresponding sequence of marginalized likelihoods from (2). In regular cases $\lim_{n\to\infty}\mathrm{p}_{jt}^{(n)}\left(\theta_j|Y_t\right)$ is a well defined posterior density function, but $\lim_{n\to\infty}M_{jt}^{(n)}=0$ and consequently $\lim_{n\to\infty}p_jM_{jt}^{(n)}/p_kM_{kt}=0$ for any model k with a fixed proper prior distribution. This is Lindley's paradox (Bartlett, 1957; Lindley, 1957) and may be paraphrased as saying that a hypothesis that assigns prior probability zero to any set of events cannot be preferred to one that assigns positive probability. "Posterior odds ratios" involving improper prior distributions for both hypotheses are especially troublesome, because they often employ convenient but arbitrary normalizing constants and therefore yield finite positive values but have no interpretation as ratios of probabilities.

An alternative approach is to regard the posterior distribution as formed from the (possibly improper) prior distribution and a subset of the data, as a prior distribution for the balance of the data. For example, Atkinson (1978) and O'Hagan (1991) propose to take $\mathrm{p}_{jt^*}\left(\theta_j\middle|Y_{t^*}\right)$ as the prior distribution, where $t^*=[\rho t]$ for some fixed $\rho\in(0,1)$, and observations $t^*+1,\mathrm{K}\ ,t$ as the sample. Then

$$M_{jt}=\int_{\Theta_j}\mathrm{p}_{jt^*}\left(\theta_j\middle|Y_t^*\right)\prod_{s=t^*+1}^{t}\mathrm{f}_{js}\left(y_s\middle|Y_{s-1},\theta_j\right)d\theta_j$$

is the marginalized likelihood, and the construction of posterior odds ratios proceeds as just described. Berger and Pericchi (1992) determine the smallest number of data points such that the posterior density is proper, form the marginalized likelihood corresponding to each subset (or a representative sample if the number of subsets is quite large) and then use the geometric mean of the resulting marginalized likelihoods in place of M_{jt}. There are several other examples of these approaches; Gelfand and Dey (1994) provide an interesting synthesis.

The way applied econometric work is actually conducted motivates an approach similar to that of Atkinson and O'Hagan. Typically the investigator has used all of the data at hand to select the model(s). This can be a sound practice, reflecting the practical decision not to undertake costly formal consideration of models whose *collective* probability clearly will be negligible compared with the model(s) selected for study. But such judgments are difficult, and even well-intentioned investigators can unwittingly tailor prior model probabilities to features of the data peculiar to the sample at hand. Pursued with premeditation, this tailoring becomes the process of "data mining" scorned by Bayesians and non-Bayesians

alike. Occasionally a portion of the sample is set aside before the process begins, but such cases are the exception and not the rule. In the predominant situation the only practical and fair arbiters between models are the predictive odds ratios. One muse use (3), with observations $u+1$ through t taken after model construction, rather than (1). The appropriate prior density is $\mathrm{p}_{ju}\left(\theta_j|Y_u\right)$. In this context we shall refer to $\mathrm{f}_{j0}\left(\theta_j\right)$ as the *protoprior density* for θ_j. The protoprior density may be improper, but the prior density $\mathrm{p}_{ju}\left(\theta_j|Y_u\right)$ typically will be proper. The question of sensitivity to the proto-prior density remains open, but both analysis and examples taken up subsequently in this paper suggest that for sufficiently large t, $\mathrm{p}_{jt}\left(\theta_j\right)$ will not be very sensitive to changes in $\mathrm{f}_{j0}\left(\theta_j\right)$.

3. The computation and updating of posterior moments

Conditional on a particular model j, most problems amount to the computation of the posterior expectation of a function of interest $\mathrm{g}_j\left(\theta_j\right)$, with the posterior density of θ_j given by (1):

$$\bar{g}_j = \mathrm{E}_t\left[\mathrm{g}_j\left(\theta_j\right)\middle|Y_t, \mathrm{f}_{j0}\left(\theta_j\right), \text{Model } j\right] = \frac{\int_{\Theta_j} \mathrm{g}_j\left(\theta_j\right)\mathrm{f}_{j0}\left(\theta_j\right)\mathrm{L}_{jt}\left(\theta_j, Y_t\right)d\theta_j}{\int_{\Theta_j} \mathrm{f}_{j0}\left(\theta_j\right)\mathrm{L}_{jt}\left(\theta_j, Y_t\right)d\theta_j}. \qquad (5)$$

Estimation, forecasting, and formal decision making each take this form, for the choice of $\mathrm{g}_j(\cdot)$ appropriate to the problem at hand. Removing the conditioning on a particular model,

$$\bar{g} = \mathrm{E}_t\left[\mathrm{g}_j\left(\theta_j\right)\middle|\left\{\text{Model } j,\, p_j,\, \mathrm{f}_{j0}\left(\theta_j\right)\right\}_{j=1}^{n}\right] = \sum_{j=1}^{n} p_j M_{jt}\bar{g}_j \Big/ \sum_{j=1}^{n} p_j M_{jt},$$

where n is the number of models under consideration, M_{jt} is the marginalized likelihood defined in (2), and p_j is the prior probability of model j. (It is implicit in this expression that the functions $\mathrm{g}_j(\cdot)$ have been chosen so that their posterior expectations pertain to the same substantive concept in the different models -- *e.g.*, the probability of a future event, an elasticity, or the value of a loss function corresponding to a particular action.)

Only in rare instances is it possible to evaluate (5) analytically, and θ_j is usually of sufficiently high dimension that deterministic computational methods like quadrature are impractical. The recent rapid development of simulation methods has made possible good numerical approximations to (5) in a wide variety of applications, however. Such methods produce a random sequence of vectors and weights $\left\{\theta_j^{(m)}, w_j^{(m)}\right\}_{m=1}^{M}$ with the property that

$$\bar{g}_j^{[M]} = \sum_{m=1}^{M} w_j^{(m)}\,\mathrm{g}_j\left(\theta_j^{(m)}\right) \Big/ \sum_{m=1}^{M} w_j^{(m)} \xrightarrow{a.s.} \bar{g}_j.$$

A principal objective of this paper is to develop methods for the practical evaluation of posterior odds ratios, building on the ability to obtain numerical approximations of $\bar{g}_j$ of this form.

We turn first, in this section, to a review of these methods (Sections 3.1 through 3.3) and an extension (Section 3.4). Their properties will be important in the development of methods for the computation of approximations to the marginalized likelihoods M_{jt} in Section 4.

3.1 Independence Monte Carlo

For certain models and prior distributions it is possible to draw θ_j directly from the posterior distribution whose probability density is given by (1). Leading examples include the univariate normal linear regression model with a normal-gamma prior, and the multinomial model with Dirichlet priors. Acceptance sampling (Geweke, 1994, Section 4.2) widens the class of models for which this is possible. For independence Monte Carlo $w_j^{(m)} \equiv 1$, and since the $\theta_j^{(m)}$ are independent, $\bar{g}_j^{[M]} \xrightarrow{a.s.} \bar{g}_j$. If $\mathrm{var}\left[\mathrm{g}_j\left(\theta_j\right)\right]$ exists and is finite then

$$M^{1/2}\left(\bar{g}_j^{[M]} - \bar{g}_j\right) \xrightarrow{d} \mathrm{N}\left(0, \sigma^2\right) \tag{6}$$

as well, with $\sigma^2 = \mathrm{var}\left[\mathrm{g}_j\left(\theta_j\right)\right]$. Since $M^{-1}\sum_{m=1}^{M}\left[\mathrm{g}_j^{(m)}\left(\theta_j\right)\right]^2 - \left[\bar{g}_j^{[M]}\right]^2 \xrightarrow{a.s.} \sigma^2$, it is straightforward to evaluate the magnitude of the approximation error $\bar{g}_j^{[M]} - \bar{g}_j$.

3.2 Importance sampling Monte Carlo

Suppose that it is not possible, or at any rate inconvenient, to draw θ_j directly from the posterior distribution. Instead let the $\theta_j^{(m)}$ be drawn from a distribution with probability density function $\mathrm{I}(\theta)$, called the importance sampling density, and let

$$w_j^{(m)} = \mathrm{w}_j\left(\theta_j^{(m)}\right) = \mathrm{p}_{jt}\left(\theta_j^{(m)} | Y_t\right) \Big/ \mathrm{I}\left(\theta_j^{(m)}\right).$$

So long as the support of $\mathrm{I}(\theta)$ includes that of $\mathrm{p}_{jt}\left(\theta_j | Y_t\right)$, $\bar{g}_j^{[M]} \xrightarrow{a.s.} \bar{g}_j$ (Geweke, 1989, Theorem 1). If in addition $\mathrm{E}\left[\mathrm{w}_j\left(\theta_j\right)\right]$ and $\mathrm{var}\left[\mathrm{g}_j\left(\theta_j\right)\right]$ exist and are finite -- *a fortiori* if $\mathrm{w}_j\left(\theta_j\right)$ is bounded above and $\mathrm{var}\left[\mathrm{g}_j\left(\theta_j\right)\right] < \infty$ -- then a limiting distribution (6) once again obtains but σ^2 is different (Geweke, 1989, Theorem 2). To obtain the value of σ^2, let $\overline{A} = \mathrm{E}_{\mathrm{I}}\left[\mathrm{w}_j\left(\theta_j\right)\mathrm{g}_j\left(\theta_j\right)\right]$ and $\overline{B} = \mathrm{E}_{\mathrm{I}}\left[\mathrm{w}_j\left(\theta_j\right)\right]$, where the subscript "I" denotes moment with respect to the importance sampling distribution with density $\mathrm{I}(\theta)$. Applying a Taylor series expansion,

$$\sigma^2 = \begin{bmatrix} 1/\overline{B} & -\overline{A}/\overline{B}^2 \end{bmatrix} \begin{bmatrix} \mathrm{var}_{\mathrm{I}}\left[\mathrm{w}_j(\theta_j)\mathrm{g}_j(\theta_j)\right] & \mathrm{cov}_{\mathrm{I}}\left[\mathrm{w}_j(\theta_j)\mathrm{g}_j(\theta_j), \mathrm{w}_j(\theta_j)\right] \\ \mathrm{cov}_{\mathrm{I}}\left[\mathrm{w}_j(\theta_j)\mathrm{g}_j(\theta_j), \mathrm{w}_j(\theta_j)\right] & \mathrm{var}_{\mathrm{I}}\left[\mathrm{w}_j(\theta_j)\right] \end{bmatrix} \begin{bmatrix} 1/\overline{B} \\ -\overline{A}/\overline{B}^2 \end{bmatrix}.$$

Substituting $\hat{\overline{A}} = M^{-1}\sum_{m=1}^{M} \mathrm{w}_j\left(\theta_j^{(m)}\right)\mathrm{g}\left(\theta_j^{(m)}\right)$ for $\overline{A}$, $M^{-1}\sum_{m=1}^{M} \mathrm{w}^2\left(\theta_j^{(m)}\right)\mathrm{g}^2\left(\theta_j^{(m)}\right) - \hat{\overline{A}}^2$ for $\mathrm{var}_{\mathrm{I}}\left[\mathrm{w}_j(\theta_j)\mathrm{g}_j(\theta_j)\right]$, etc., the central limit theorem becomes operational.

For importance sampling Monte Carlo to be effective $\mathrm{I}(\theta_j)$ must approximate $\mathrm{p}_{jt}\left(\theta_j^{(m)}|Y_t\right)$ well in the appropriate way. If $\mathrm{E}\left[\mathrm{w}_j(\theta_j)\right]$ is not finite (implying $\mathrm{w}_j(\theta_j)$ is unbounded) then not only is there no limiting distribution (6), but convergence is usually impractically slow. If $\mathrm{E}\left[\mathrm{w}_j(\theta_j)\right] = \mathrm{E}_{\mathrm{I}}\left[\mathrm{w}_j^2(\theta_j)\right] < \infty$ but the right tail of the distribution of $\mathrm{w}_j^2(\theta_j)$ under the importance sampling distribution is sufficiently important then convergence can still be very slow even in simple problems; see Geweke (1989) for examples. In general and as a practical matter, importance sampling Monte Carlo is effective to the extent that $\mathrm{w}_j(\theta_j)$ can be bounded above by a constant not too large relative to $\mathrm{E}\left[\mathrm{w}_j(\theta_j)\right]$.

3.3 Markov chain Monte Carlo

Following a line of research that began with Metropolis *et al.* (1954), several investigators have recently constructed algorithms in which $\left\{\theta_j^{(m)}\right\}_{m=1}^{\infty}$ is a realization of a continuous state Markov chain, with the properties

$$\theta_j^{(m)} \xrightarrow{d} \mathrm{P}_{jt}, \quad \overline{g}_j^{[M]} \xrightarrow{a.s.} \overline{g}_j, \tag{7}$$

where P_{jt} is the distribution corresponding to (1). (Thus, $\mathrm{w}_j(\theta_j) \equiv 1$.) One example is the Gibbs sampling algorithm developed by Geman and Geman (1984), Gelfand and Smith (1990), and others. Casella and George (1992) provide an introductory exposition, and there are examples in Sections 5.2 and 6.1. Another example is the Metropolis chain proposed by Metropolis *et al.* (1954) and extended by Hastings (1970) and others. Chib and Greenberg (1994) provide a good introduction, and there is an example in Section 5.1. An extension of considerable importance to econometrics is the data augmentation algorithm of Tanner and Wong (1987). Their essential contribution is to note that in a subjective Bayesian approach parameters and latent variables are inherently symmetric, and therefore the Gibbs sampling algorithm obviates the need to integrate explicitly over the distribution of the latent variables. An example is provided in Section 5.2.

Conditions under which (7) obtains for any $\theta_j^{(1)} \in \Theta_j$ are rather general, and include essentially all conventional econometric models: Tierney (1991) and references cited therein provide weak sufficient conditions; Roberts and Smith (1992) present conditions that are stronger, easier to verify, and usually obtain in econometric applications. Even though these conditions are satisfied, it is conventional to discard some initial simulations to mitigate sensitivity to initial conditions. These authors also discuss conditions under which there exists a central limit theorem of the form (6); see also Geyer (1992). Conditions under which σ^2 can be approximated consistently in M are more elusive, and for useful discussions the reader is referred to Gelman and Rubin (1992) and the comments that follow. Growing experience with these methods suggests that in econometric models $\left\{g_j\left(\theta_j^{(m)}\right)\right\}$ generally behaves like a stationary stochastic process, with a spectral density function which we shall denote $S(\lambda)$. Standard frequentist time series analysis (e.g., Hannan, 1970, pp. 207-210) then yields $\sigma^2 = S(0)$, and well-established procedures may be used to approximate S(0); Geweke (1992, Section 3) provides details.

In most applications the bulk of the computing time is devoted to drawing the $\theta_j^{(m)}$. The numerical efficiency of Markov chain Monte Carlo methods depends on computation time for each iteration, and on the degree of serial correlation in $\left\{g_j\left(\theta_j^{(m)}\right)\right\}$. Very strong positive serial correlation implies $S(0) >> \mathrm{var}\left[g\left(\theta_j^{(m)}\right)\right]$, so that relatively many more iterations will be required than if independence Monte Carlo sampling had been possible.

3.4 Importance sampling Markov chain Monte Carlo (updating)

Importance sampling and Markov chain Monte Carlo can be combined. To motivate the combination, consider the problem facing a Bayesian econometrician wishing to update $E_t\left[g_j\left(\theta_j\right)\right]$ in real time. If the econometrician had applied one of the foregoing methods he/she could, of course, simply perform the same analysis with the new posterior density $p_{j,t+1}\left(\theta_j|Y_{t+1}\right)$ in lieu of the old one. In the case of importance sampling this requires M new evaluations of the posterior density and the importance sampling density, with computation time generally proportional to $t+1$. In the case of Markov chain Monte Carlo this requires a complete generation of a new set of $\theta_j^{(m)}$ from the updated Markov chain.

Alternatively, the econometrician can regard (1) as an importance sampling density for the updated posterior density $p_{j,t+1}\left(\theta_j|Y_{t+1}\right)$. The appropriate weight function is simply

$$L_{j,t+1}\left(\theta_j, Y_{t+1}\right) / L_{j,t}\left(\theta_j, Y_t\right) = f_{j,t+1}\left(y_{t+1}|Y_t, \theta_j\right).$$

In virtually all applications this function is bounded above and it is then straightforward to show that

$$\sum_{m=1}^{M} \mathrm{w}^*\left(\theta_j^{(m)}\right)\mathrm{g}\left(\theta_j^{(m)}\right) \Big/ \sum_{m=1}^{M} \mathrm{w}^*\left(\theta_j^{(m)}\right) \xrightarrow{a.s.} \bar{g}_j$$

where the $\theta_j^{(m)}$ are drawn in any of the ways discussed above and

$$\mathrm{w}_j^*\left(\theta_j\right) = \mathrm{w}_j\left(\theta_j\right)\mathrm{f}_{j,t+1}\left(y_{t+1}\middle|Y_t,\theta_j\right).$$

Since this approximation avoids the need to regenerate a new sequence $\left\{\mathrm{g}\left(\theta_j^{(m)}\right)\right\}_{m=1}^{M}$ it is generally quite fast -- up to $t+1$ times faster in the limiting but not atypical case in which essentially all computing time is spent in evaluation of the posterior density $\mathrm{p}_{j,t+1}\left(\theta_j\middle|Y_{t+1}\right)$ or in drawing the $\theta_j^{(m)}$ from the Markov chain. The only essential new computations are the evaluations of $\mathrm{f}_{j,t+1}\left(y_{t+1}\middle|Y_t,\theta_j\right)$.

The computational efficiency of this procedure depends on the variation in $\mathrm{w}_j^*\left(\theta_j\right)$ with respect to the posterior distribution of θ_j based on t observations. For most conditional densities in econometric models this variation will be greater when y_{t+1} is an outlier (*i.e.*, $\mathrm{f}_{j,t+1}\left(y_{t+1}\middle|Y_t,\theta_j\right)$ is smaller) than when it is not (*i.e.*, $\mathrm{f}_{j,t+1}\left(y_{t+1}\middle|Y_t,\theta_j\right)$ is larger).

This procedure may be extended in an obvious way to several observations, with

$$\mathrm{w}_j^*\left(\theta_j\right) = \mathrm{w}_j\left(\theta_j\right)\prod_{s=1}^{r}\mathrm{f}_{j,t+s}\left(y_{t+s}\middle|Y_{t+s-1},\theta_j\right).$$

It is limited by the fact that as r increases the maximum value of $\prod_{s=1}^{r}\mathrm{f}_{j,t+s}\left(y_{t+s}\middle|Y_{t+s-1},\theta_j\right)$ is increasing relative to its posterior mean. When r is sufficiently great it will be more efficient to repeat the original algorithm using the sample with $t+r$ observations in lieu of the one with t observations.

The numerical accuracy of this procedure may be assessed as described in Section 3.2 if the original computational procedure involved simple or importance sampling Monte Carlo. In the case of Markov chain Monte Carlo, let $\mathrm{S}_w(\lambda)$ denote the spectral density of $\left\{\mathrm{w}_j\left(\theta_j^{(m)}\right)\right\}$, $\mathrm{S}_{wg}(\lambda)$ the spectral density of $\left\{\mathrm{w}_j\left(\theta_j^{(m)}\right)\mathrm{g}_j\left(\theta_j^{(m)}\right)\right\}$, and $\mathrm{S}_{wwg}(\lambda)$ the cross spectral density of $\left\{\mathrm{w}_j\left(\theta_j^{(m)}\right)\right\}$ and $\left\{\mathrm{w}_j\left(\theta_j^{(m)}\right)\mathrm{g}_j\left(\theta_j^{(m)}\right)\right\}$. As before define $\bar{A} = \mathrm{E}_{\mathrm{I}}\left[\mathrm{w}_j\left(\theta_j\right)\mathrm{g}_j\left(\theta_j\right)\right]$ and $\bar{B} = \mathrm{E}_{\mathrm{I}}\left[\mathrm{w}_j\left(\theta_j\right)\right]$. Then by the same arguments that lead to $\sigma^2 = \mathrm{S}(0)$ in Section 3.3,

$$\sigma^2 = \left[1/\bar{B} \quad -\bar{A}/\bar{B}^2\right]\begin{bmatrix} \mathrm{S}_{wg}(0) & \mathrm{S}_{wwg}(0) \\ \mathrm{S}_{wwg}^*(0) & \mathrm{S}_w(0) \end{bmatrix}\begin{bmatrix} 1/\bar{B} \\ -\bar{A}/\bar{B}^2 \end{bmatrix}.$$

Replacing each constituent of the right hand side with its consistent (in M) estimator yields an operational approximation to σ^2.

4. The practice of model comparison

The composition of the posterior odds ratio and its decomposition into linked predictive likelihoods, and simulation-based methods for the approximation of posterior moments, taken together suggest a new technology for model comparison and the public reporting of the results of Bayesian inference. In this section we outline the important aspects of these new procedures.

4.1 Systematic comparison of marginalized likelihoods

From expressions (3) and (4),

$$\hat{p}_{ju}^{t} = \prod_{l=1}^{q} \hat{p}_{js_{l-1}}^{s_l},$$

where

$$\hat{p}_{js_{l-1}}^{s_l} = \frac{\int_{\theta_j} \mathrm{p}_{js_0}(\theta_j) \prod_{s=s_0+1}^{s_l} \mathrm{f}_{js}(y_s | Y_{s-1}, \theta_j) d\theta_j}{\int_{\theta_j} \mathrm{p}_{js_0}(\theta_j) \prod_{s=s_0+1}^{s_{l-1}} \mathrm{f}_{js}(y_s | Y_{s-1}, \theta_j) d\theta_j} \qquad (8)$$

and $0 \le u = s_0 < s_1 < \mathrm{K} < s_q = t$. (If $u = 0$ then $\hat{p}_{ju}^{t} = M_{jt}$.) Expression (8) is precisely in the form of (5), with the posterior density kernel composed of the prior density $\mathrm{p}_{js_0}(\theta_j)$ and the likelihood function for the first s_{l-1} observations, and the function of interest is the likelihood function for observations $s_{l-1}+1$ through s_l. It is immediately evident that the predictive likelihood, and therefore the marginalized likelihood, can be evaluated using one of the simulation methods of Section 3, with appropriate definitions of the posteriors and functions of interest. Special methods (*e.g.*, Spiegelhalter and Smith, 1982; Newton and Raftery, 1994) are not required.

The foregoing arguments of this paper suggest how one might use (8) in practice. As discussed in Section 2.2, the choice of u may depend on the way the model has been constructed. In many instances it may be desirable to choose a date corresponding to the creation of the model. In any event, the interpretation of $\hat{p}_{ju}^{t}$ is clear for any stated u: observations through u are treated as part of a training sample that enters the prior, and subsequent observations form the basis for model comparison.

The complete decomposition $s_l - s_{l-1} = 1$ is attractive as a diagnostic for reasons discussed in Section 2.1. The computational procedures developed in Section 3.4 show that such a decomposition need not be computationally burdensome. Consider the problem of forming $\hat{p}_{js}^{s+1}, \hat{p}_{j,s+1}^{s+2}, \mathrm{K}$ given a simulated Monte Carlo sample of $\theta_j^{(m)}$ from the posterior density $\mathrm{p}_{js}(\theta_j | Y_s)$. The numerical approximation of $\hat{p}_{js}^{s+1}$ is the weighted average

$$\sum_{m=1}^{M} \mathrm{w}_j\left(\theta_j^{(m)}\right) \mathrm{f}_{j,s+1}\left(y_{s+1}\middle|Y_s, \theta_j^{(m)}\right) \Big/ \sum_{m=1}^{M} \mathrm{w}_j\left(\theta_j^{(m)}\right).$$

But the $\mathrm{f}_{j,s+1}\left(y_{s+1}\middle|Y_s, \theta_j^{(m)}\right)$ are precisely the weights required to transform the simulated Monte Carlo sample from $\mathrm{p}_{js}\left(\theta_j|Y_s\right)$ to a simulated Monte Carlo sample from $\mathrm{p}_{j,s+1}\left(\theta_j|Y_{s+1}\right)$, as described in Section 3.4. This establishes a recursion for the computation of the $\hat{p}_{js}^{s+1}, \hat{p}_{j,s+1}^{s+2}, \mathrm{K}$ that does not require one to recompute a Monte Carlo sample from the entire posterior distribution.

This recursion remains practical so long as the importance sampling weights

$$\prod_{u=1}^{r} \mathrm{f}_{j,s+u}\left(y_{s+u}\middle|Y_{s+u-1}, \theta_j^{(m)}\right) \quad (m = 1, \mathrm{K}\ , M)$$

remain well behaved. As r increases, however, $\mathrm{p}_{js}\left(\theta_j|Y_s\right)$ becomes a poorer importance sampling density for $\mathrm{p}_{j,s+r}\left(\theta_j|Y_{s+r}\right)$. The case of a shift in regime sometime between s and $s+r$ aside, the problem is not violation of the conditions in the first paragraph of Section 3.2, but rather that excessive dispersion in $\mathrm{p}_{js}\left(\theta_j|Y_s\right)$ relative to $\mathrm{p}_{j,s+r}\left(\theta_j|Y_{s+r}\right)$ eventually makes it very inefficient. As this occurs it eventually becomes more efficient to collect the new r observations into a full new posterior density $\mathrm{p}_{j,s+r}\left(\theta_j|Y_{s+r}\right)$, construct $\left\{\theta_j^{(m)}\right\}_{m=1}^{M}$ from this posterior, and begin to compute $\hat{p}_{j,s+r}^{s+r+1}, \hat{p}_{j,s+r+1}^{s+r+2}, \mathrm{K}$. The next subsection presents some results bearing on how this might be done; the reader not concerned with details of computation can proceed to Section 4.3 without loss of continuity.

4.2 Some comparisons of computational efficiency

For reasons just discussed, suppose we wish to approximate the expectation of

$$\prod_{u=1}^{r} \mathrm{f}_{j,s+u}\left(y_{s+u}\middle|Y_{s+u-1}, \theta_j^{(m)}\right) \tag{9}$$

against the posterior density $\mathrm{p}_{js}\left(\theta_j|Y_s\right)$ for θ_j. This can be done either by treating (9) as a function of interest directly, or by forming numerical approximations to each $\mathrm{f}_{j,s+u}\left(y_{s+u}\middle|Y_{s+u-1}, \theta_j^{(m)}\right)$ separately and then forming the product. Computational efficiency argues for the latter, not the former.

To see why, let $X_u^{(m)} = \mathrm{f}_{j,s+u}\left(y_{s+u}\middle|Y_{s+u-1}, \theta_j^{(m)}\right)$, $\mu_u = \mathrm{E}\left(X_u^{(m)}\right)$, $\sigma_u^2 = \mathrm{var}\left(X_u^{(m)}\right)$, where the moments are with respect to the posterior density $\mathrm{p}_{js}\left(\theta_j|Y_s\right)$. (To simplify the notation we take up the case $\mathrm{w}_j\left(\theta_j\right) = 1$ and drop the $\mathrm{w}_j\left(\theta_j\right)$. Analysis with varying weights proceeds in the same way.) Since $X_u^{(m)} \geq 0$ and $X_u^{(m)}$ is bounded above, all moments of $X_u^{(m)}$ must exist. For the second procedure (approximate each function first, then take the product)

$$\mathrm{MSE}\left\{\prod_{u=1}^{r} M^{-1} \sum_{m=1}^{M} X_u^{(m)}\right\} = \left(\prod_{u=1}^{r} \mu_u^2\right) \sum_{u=1}^{r} \sigma_u^2 / M\mu_u^2 + \mathrm{o}\left(M^{-1}\right). \tag{10}$$

For the first procedure (approximate the product directly)

$$\mathrm{MSE}\left\{M^{-1}\sum_{m=1}^{M}\left(\prod_{u=1}^{r}X_u^{(m)}\right)\right\}=M^{-1}\left[\prod_{u=1}^{r}\left(\mu_u^2+\sigma_u^2\right)-\prod_{u=1}^{r}\mu_u^2\right]$$
$$>M^{-1}\sum_{u=1}^{r}\sigma_u^2\prod_{j\neq u}\mu_j^2+\mathrm{o}\left(M^{-1}\right)=M^{-1}\left(\prod_{u=1}^{r}\mu_u^2\right)\sum_{u=1}^{r}\sigma_u^2/\mu_u^2+\mathrm{o}\left(M^{-1}\right).$$

To appreciate the magnitudes involved consider a simple case that can be treated analytically: $y_t \overset{IID}{\sim}\mathrm{N}\left(\theta,\mathrm{I}_k\right)$, and suppose sample size is sufficiently large that the prior distribution can be neglected. If sample size is t and $k=1$,

$$\mathrm{f}_{j,t+1}\left(y_{t+1}|Y_t,\theta\right)=(2\pi)^{-1/2}\exp\left[-\left(y_{t+1}-\theta\right)^2/2\right]. \tag{11}$$

Suppose further that the posterior distribution of θ following t observations is $\mathrm{N}\left(0,t^{-1}\right)$. For independence Monte Carlo the first moment for the function (11) of interest is

$$\int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp\left[-\left(y_{t+1}-\theta\right)^2/2\right](2\pi)^{-1/2}t^{1/2}\exp\left(-\theta^2/2t^{-1}\right)d\theta$$
$$=(2\pi)^{-1/2}\left[t/(t+1)\right]^{1/2}\exp\left[-ty_{t+1}^2/2(1+t)\right],$$

and the second moment is

$$(2\pi)^{-1}\left[t/(t+2)\right]^{1/2}\exp\left[-ty_{t+1}^2/(t+2)\right].$$

The mean square error of numerical approximation given M replications is

$$(2\pi M)^{-1}\left\{\left(\frac{t}{t+2}\right)^{1/2}\exp\left(\frac{-ty_{t+1}^2}{t+2}\right)-\left(\frac{t}{t+1}\right)\exp\left(\frac{-ty_{t+1}^2}{t+1}\right)\right\}.$$

For $k>1$ or $r>1$, $\prod_{s=1}^{r}\mathrm{f}_{j,t+s}\left(y_{t+s}|Y_{t+s-1},\theta\right)$ factors into rk independent components each of the same form as (11), and the MSE is

$$M^{-1}(2\pi)^{-rk}\left\{\left(\frac{t}{t+2}\right)^{rk/2}\exp\left[\frac{-t}{t+2}\sum_{i=1}^{k}\sum_{s=1}^{r}y_{i,t+s}^2\right]-\left(\frac{t}{t+1}\right)^{rk}\exp\left[\frac{-t}{t+1}\sum_{i=1}^{k}\sum_{s=1}^{r}y_{i,t+s}^2\right]\right\}. \tag{12}$$

Suppose that instead one approximates each of the r factors and then takes the product. Applying (10) one obtains the MSE

$$M^{-1}(2\pi)^{-rk}\left(\frac{t}{t+1}\right)^{rk}\left\{\frac{(t+1)^k}{\left[t(t+2)\right]^{k/2}}\sum_{u=t+1}^{t+r}\exp\left[\frac{-t}{t+1}\sum_{i=1}^{k}\sum_{s\neq u}y_{is}^2-\frac{t}{t+2}\sum_{i=1}^{k}y_{iu}^2\right]\right.$$
$$\left.-\sum_{u=t+1}^{t+r}\exp\left[\frac{-t}{t+1}\sum_{i=1}^{k}\sum_{u=t+1}^{t+r}y_{iu}^2\right]\right\}. \tag{13}$$

Comparing (12) and (13) obviously involves the data $y_{t+1},\mathrm{K},y_{t+r}$. Maintaining the assumption $y_t \overset{IID}{\sim}\mathrm{N}\left(0,\mathrm{I}_k\right)$, expectations over y_t yield mean values of

$$(2\pi)^{-rk}\left[\left(\frac{t}{3t+2}\right)^{rk/2}-\frac{t^{rk}}{\left[(t+1)(3t+1)\right]^{rk/2}}\right] \tag{14}$$

and

$$(2\pi)^{-rk}r\left(\frac{t}{t+1}\right)^{rk}\left(\frac{t+1}{3t+1}\right)^{(r-1)k/2}\left[\frac{(t+1)^k}{\left[t(3t+2)\right]^{k/2}}-\left(\frac{t+1}{3t+1}\right)^{k/2}\right] \tag{15}$$

for (12) and (13), respectively. The ratio of (14) to (15), for some alternative values of t and r with $k = 9$ are provided in Table 1. For $r = t$, numerical approximation of the entire product incurs MSE about six times greater than the product of the numerically approximated predictive likelihoods.

The expressions (12)-(13) are predicated on the assumption that the posterior distribution of θ is Gaussian, which may be reasonable for large sample sizes even when the model itself is not Gaussian. The assumption that the data are Gaussian, made in moving from (12)-(13) to (14)-(15), is not so general. For example, if the distribution of the y_{it} is Student-*t* then expectations of (12) and (13) are not even finite. (The expressions involve the moment generating function of the central *F* distribution, which does not exist.) Therefore the values in Table 1 should be viewed as quite conservative. This fact is borne out in the examples taken up in Section 6.

4.3 Econometric tests and public reporting

These methods, and continued advances in computation and communication, have implications for public reporting by Bayesians. Predictive likelihoods can be calculated routinely and ought to be reported in published work. Beyond this, computational devices for the integration of public reporting and private subjective priors are clearly at hand. From the sampled $\theta_j^{(m)}$, and the investigator's prior distribution, any other investigator can use importance sampling Monte Carlo (Section 3.4) to

(i) impose his/her subjective priors;
(ii) investigate the sensitivity of posterior moments to prior distributions;
(iii) evaluate the posterior expectation of other functions of interest not considered by the original investigator; or,
(iv) update the original posterior distribution with new observations.

Observe that the theory of Monte Carlo approximation error outlined in Section 3.2 argues that the ratio of a client's prior density kernel to the prior density kernel used in public reporting should be bounded. In many instances this may commend the use of a flat or similarly diffuse prior in public reporting even though no client entertains such a prior. In sampling methods involving data augmentation, it may often be efficient to report a subset of the sampled values of the latent variables to facilitate updating the posterior. (Section 5.2 provides such an example of data augmentation.) The sampled $\theta_j^{(m)}$ provide a set of sufficient statistics for the numerical approximation of all posterior moments in question. Their computation may require specialized software and substantial time, but once these computations have been completed further analysis along the lines of (i) - (iv) can be done rapidly.

These considerations argue that investigators should provide the $\theta_j^{(m)}$ and corresponding evaluations of their prior density (if it is not flat), for $\theta_j^{(m)}$ drawn from the entire posterior and for enough subsamples to permit efficient numerical approximations. Additionally they should provide software in a standard low-level language for the evaluation of $f_{js}\left(y_s \middle| Y_{s-1}, \theta_j\right)$. These procedures can lead to large files, but costs of storage and remote access continue to decline.

5. An example: GARCH and stochastic volatility models

Models in which the volatility of asset returns varies smoothly over time have received considerable attention in recent years. (For a survey of several approaches see Bollerslev, Chou and Kroner (1992).) Persistent but changing volatility is an evident characteristic of returns data. Since the conditional distribution of returns is relevant in the theory of portfolio allocation, proper treatment of volatility is important. Time-varying volatility also affects the properties of real growth and business cycle models.

The earliest model of time varying volatility is the autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982). This was extended to the generalized ARCH (GARCH) model by Bollerslev (1986). Since then many variants of ARCH models have appeared. The distinguishing characteristic of these models is that the conditional variance of the return is a deterministic function of past conditional variances and past values of the return itself. GARCH models exhibit both time-varying volatility and leptokurtic unconditional distributions, but the two cannot be separated: these models cannot account for leptokurtosis without introducing time-varying volatility.

Stochastic volatility models have been examined by a series of investigators beginning with Taylor (1986). Promising Bayesian methods, used here, have been developed by Jacquier, Polson and Rossi (1994). In these models the conditional variance of the return is a stochastic function of its own past values but is unaffected by past returns themselves. Like GARCH models they account for time-varying volatility and leptokurtosis, but unlike GARCH models it is possible to have excess kurtosis without heteroscedasticity.

In this section we compare these two models, using the methods set forth in the paper and a time series of 3,010 daily closing observations of the U.S./Canadian exchange rate.

5.1 The GARCH model

The GARCH model of time-varying volatility may be expressed

$$y_t = \beta' \mathbf{x}_t + h_t^{1/2} \varepsilon_t$$
$$h_t = \alpha + \sum_{s=1}^{q} \gamma_s \varepsilon_{t-s}^2 + \sum_{j=1}^{p} \delta_j h_{t-j} \qquad (16)$$
$$\varepsilon_t \sim \text{IIDN}(0,1)$$

Here, y_t is the observed return at time t; $\mathbf{x}_t$ is a vector of covariates and β is the corresponding vector of coefficients; h_t is the conditional variance at time t; $\alpha > 0$, $\gamma_s \geq 0$ $(s = 1, \text{K}, q)$, $\delta_j \geq 0$ $(j = 1, \text{K}, p)$. The vector of covariates is typically deterministic, including a constant term and perhaps indicator variables for calendar effects on the mean of y_t.

For the comparisons taken up here we use only the GARCH (1,1) model, which is (16) with $p = q = 1$. (Henceforth, we omit the subscripts on γ_1 and δ_1.) The GARCH (1,1) specification has proven attractive for models of returns. It typically dominates other GARCH models using the Akaike or Schwarz Bayesian information criteria (Bollerslev, Chou and Kroner, 1992). Following the GARCH literature we treat h_1 as a known constant. Then, the likelihood function is

$$\text{L}_u(\beta, \alpha, \gamma, \delta | Y_u) = \prod_{s=1}^{u} h_s^{1/2} \exp\left[-(y_s - \mathbf{x}_s'\beta)^2 / 2h_s\right] \qquad (17)$$

where h_s is computed recursively from (16). The predictive density, through observation t, is

$$(2\pi)^{-(t-u)/2} \prod_{s=u+1}^{t} h_s^{1/2} \exp\left[-(y_s - \mathbf{x}_s'\beta)^2 / 2h_s\right].$$

For expressing prior distributions as well as for carrying out the computations it proves useful to work with $a = \log(\alpha)$ rather than α. With this reparameterization the functional form of the prior distribution used in this work is

$$a \sim \text{N}(\underline{a}, \underline{s}_a^2);$$
$$\beta \sim \text{N}(\underline{\beta}, \underline{S}_\beta); \qquad (18)$$
$$\pi(\gamma, \delta) = 2 \ (\gamma \geq 0, \delta \geq 0, \ \gamma + \delta < 1);$$

and the distributions are independent. Restriction of γ and δ to the unit simplex is equivalent to the statement that the variance process is stationary. Choices of the parameters of the prior distributions and sensitivity of the results to these choices are taken up in Section 5.3.

To perform the computations we construct a Metropolis independence chain to produce a sequence of parameters whose unconditional limiting distribution is the posterior distribution. Let $\theta' = (\beta', a, \gamma, \delta)$, and let $\text{p}_{1t}(\theta | Y_t)$ denote the posterior distribution at time t. The kernel of this distribution is the product of (17) and the three prior density kernels

in (18). The mode of the log posterior kernel is easily found using analytical expressions for the gradient and Hessian and a standard Newton-Raphson algorithm. Denote the mode by $\hat{\theta}$, and the Hessian at the mode by $\mathbf{H}$. Let $\mathrm{J}(\cdot;\mu, \mathbf{V}, \nu)$ denote the kernel density of a multivariate Student-*t* distribution with location vector μ, scale matrix $\mathbf{V}$, and ν degrees of freedom. For the choices $\mu = \hat{\theta}, \mathbf{V} = -(1.2)^2\mathbf{H}^{-1}, \nu = 5$, the ratio $\mathrm{p}_{1t}\left(\theta|Y_t\right)\big/\mathrm{J}\left(\theta;\mu,\mathbf{V},\nu\right)$ is bounded above (as indicated by a Newton-Raphson algorithm).

This multivariate Student-*t* distribution forms a proposal distribution for an independence Metropolis algorithm as follows. At step m, generate a candidate θ^* from $\mathrm{J}(\cdot;\mu, \mathbf{V}, \nu)$. With probability

$$p = \min\left\{\frac{\mathrm{p}_{1t}\left(\theta^*|Y_t\right)\big/\mathrm{J}\left(\theta^*;\mu,\mathbf{V},\nu\right)}{\mathrm{p}_{1t}\left(\theta^{(m-1)}|Y_t\right)\big/\mathrm{J}\left(\theta^{(m-1)};\mu,\mathbf{V},\nu\right)},\ 1\right\},$$

$\theta^{(m)} = \theta^*$; and with probability $1-p$, $\theta^{(m)} = \theta^{(m-1)}$. For this proposal distribution, about half the candidate parameter vectors were accepted in the work discussed below.

Given the sequence $\left\{\theta^{(m)}\right\}$ formed in this way, marginalized likelihoods were computed by the method of linked predictive likelihoods described in Sections 2 and 3. Over the next 20 observations, $M = 2000$ iterations of the independence Metropolis chain produce numerical approximations of the predictive likelihood whose numerical standard error is typically about 0.5% of the predictive likelihood, and never greater than 1%. One can therefore obtain reliable marginalized likelihoods using (4) with intervals of 20 observations between the s_1. Computation time ranged from just under 75 seconds for the smallest sample (757 observations) to just over 5 minutes for the largest sample (2962 observations), on a Sun 10/51.

5.2 The stochastic volatility model

The stochastic volatility model taken up by Jacquier, Polson and Rossi (1994) is

$$\begin{aligned} y_t &= \beta'\mathbf{x}_t + \varepsilon_t, \quad \varepsilon_t = h_t^{1/2}u_t, \\ \log h_t &= \alpha + \delta\log h_{t-1} + \sigma_v v_t, \\ \begin{pmatrix} u_t \\ v_t \end{pmatrix} &\overset{IID}{\sim} \mathrm{N}(0, \mathbf{I}_2), \end{aligned} \tag{19}$$

where $|\delta| < 1$ and $\sigma_v > 0$. Following Jacquier, Polson and Rossi we do not condition on h_1 but rather regard h_1 as a random variable drawn from its unconditional distribution $\mathrm{N}\left(\alpha/(1-\delta),\ \sigma_v^2/(1-\delta^2)\right)$. Then,

$$\mathrm{L}_u\left(\beta,\alpha,\delta,\sigma_v|Y_u\right) = \int_0^\infty \mathrm{L} \int_0^\infty \mathrm{L}_u^*\left(\beta,\alpha,\delta,\sigma_v,h_1,\mathrm{K}\ ,h_u|Y_u\right)dh_1\mathrm{L}\ dh_u$$

where

$$\mathrm{L}_u^*(\beta,\alpha,\delta,\sigma_v,h_1,\mathrm{K}\ ,h_u|Y_u)=$$
$$\prod_{s=1}^{u}h_s^{-3/2}\exp\Big(-\sum_{s=1}^{u}\varepsilon_s^2/2h_s\Big)\exp\Big[-\sum_{s=2}^{u}\big(\log h_s-\alpha-\delta\log h_{s-1}\big)^2/2\sigma_v^2\Big] \quad (20)$$
$$\cdot\exp\Big\{\big[\log h_1-\alpha/(1-\delta)\big]^2/\big[\sigma_v^2/(1-\delta^2)\big]\Big\}.$$

The prior distributions for β and σ_v are of the forms

$$\beta\sim \mathrm{N}\big(\underline{\beta},\underline{S}_\beta\big) \quad (21a)$$

and

$$\underline{\nu}_v\,\underline{s}_v^2/\sigma_v^2\sim\chi^2(\underline{\nu}_v), \quad (21b)$$

respectively. The prior distribution of $(\alpha,\delta)'$ is bivariate normal, induced by independent normal prior distributions on the persistence parameter δ,

$$\delta\sim\mathrm{N}\big(\underline{\delta},\underline{s}_\delta^2\big)$$

and the unconditional mean of $\log h_t$,

$$\alpha/(1-\delta)\sim\mathrm{N}\big(\underline{h},\,\underline{s}_h^2\big).$$

A linearization of $\alpha/(1-\delta)$ yields the corresponding bivariate normal prior distribution,

$$\begin{pmatrix}\alpha\\ \delta\end{pmatrix}\sim\mathrm{N}\left(\begin{bmatrix}\underline{h}(1-\underline{\delta})\\ \underline{\delta}\end{bmatrix},\ \begin{bmatrix}\underline{s}_h^2(1-\underline{\delta})^2+\underline{h}^2\underline{s}_\delta^2 & -\underline{h}\underline{s}_\delta^2\\ -\underline{h}\underline{s}_\delta^2 & \underline{s}_\delta^2\end{bmatrix}\right). \quad (21c)$$

To perform the computations we construct a Gibbs sampling - data augmentation algorithm. To describe this procedure, let $\theta'=(\beta',\alpha,\delta,\sigma_v)$ and $\mathbf{h}'=(h_1,\mathrm{K}\ ,h_u)$, and note that for any function of interest $\mathrm{g}(\theta,\mathbf{h})$ we can write

$$\mathrm{E}[\mathrm{g}(\theta,\mathbf{h})]=\frac{\int_\Theta \mathrm{g}(\theta)\mathrm{L}_u(\theta|Y_u)\pi(\theta)d\theta}{\int_\Theta \mathrm{L}_u(\theta|Y_u)\pi(\theta)d\theta}=\frac{\int_\Theta\int_H \mathrm{g}(\theta)\mathrm{L}_u^*(\theta,\mathbf{h}|Y_u)\pi(\theta)d\mathbf{h}d\theta}{\int_\Theta\int_H \mathrm{L}_u^*(\theta,\mathbf{h}|Y_u)\pi(\theta)d\mathbf{h}d\theta},$$

where $\pi(\theta)$ is the prior distribution constructed from (21). Thus, the latent variables $h_1,\mathrm{K}\ ,h_u$ are symmetric to the parameters θ in the Bayesian inference problem.

In the Gibbs sampling algorithm, successive subvectors of parameters and latent variables are drawn conditional on the remaining parameters and latent variables. The conditions of Roberts and Smith (1992) for convergence of this process to the posterior distribution are satisfied in this model. For the parameter vector θ the Gibbs sampling algorithm employed here is the same as that used by Jacquier, Polson and Rossi (1994): the posterior distribution of β conditional on (α,δ), σ_v and $\mathbf{h}$ is normal; the posterior distribution of (α,δ) conditional on β, σ_v and $\mathbf{h}$ is normal up to the last term of (20) which may be accommodated by acceptance sampling; and the distribution of σ_v conditional on β, (α,δ) and $\mathbf{h}$ is inverted gamma.

The treatment of $\mathbf{h}$ differs from that of Jacquier, Polson and Rossi (1994). The posterior distribution of h_s $(1<s<u)$, conditional on $\{h_r, r\neq s\}$ and θ has density kernel

$$h_s^{-3/2}\exp\left(-\varepsilon_s^2/2h_s\right)\exp\left[-\left(\log h_s-\mu_s\right)^2/2\sigma^2\right]$$

where

$$\varepsilon_s = y_s - \mathbf{x}_s'\beta,\quad \mu_s = \frac{\alpha(1-\delta)+\delta\left(\log h_{s-1}+\log h_{s+1}\right)}{1+\delta^2},\quad \sigma^2 = \frac{\sigma_v^2}{1+\delta^2}.$$

The posterior conditional density kernel for $H_s = \log h_s$ is

$$\exp\left[-\left(H_s-\mu_s^*\right)/2\sigma^2\right]\exp\left[-\varepsilon_s^2/2\exp\left(H_s\right)\right],$$

where $\mu_s^* = \mu_s - .5\sigma^2$. One can draw efficiently from this distribution using acceptance sampling, employing a source $\mathrm{N}\left(\lambda,\sigma^2\right)$ distribution with λ chosen optimally as described in Geweke (1994, Section 3.2). For $H_1 = \log h_1$ the conditional posterior density kernel is

$$\exp\left[-\left(H_1-\mu_1^*\right)^2/2\sigma_v^2\right]\exp\left[-\varepsilon_1^2/2\exp\left(H_1\right)\right]$$

where $\mu_1^* = \alpha + \delta H_2 - .5\sigma_v^2$. There is a symmetric expression for $H_u = \log h_u$.

The predictive density for observations $u+1,\mathrm{K}\ ,t$, given observations $1,\mathrm{K}\ ,u$, is

$$(2\pi)^{-(t-u)/2}\int_0^\infty \mathrm{L}\int_0^\infty \prod_{s=u+1}^t h_s^{-3/2}\exp\left(\sum_{s=u+1}^t -\varepsilon_s^2/2h_s^{1/2}\right)\mathrm{p}^*\left(h_{u+1},\mathrm{K}\ ,h_t\middle|h_1,\mathrm{K}\ h_u,\theta\right)dh_{u+1}\mathrm{L}\ dh_t$$

where $\varepsilon_s = y_s - \mathbf{x}_s'\beta$ and $\mathrm{p}^*\left(\cdot\middle|\cdot\right)$ is the distribution of $h_{u+1},\mathrm{K}\ ,h_t$ conditional on the latent $h_1,\mathrm{K}\ ,h_u$ and the specified value of the parameter vector. The simulated values of the predictive likelihood may therefore be formed as follows. For each sampled $\left(\theta^{(m)},\mathbf{h}^{(m)}\right)$, draw R sets of $h_{u+1},\mathrm{K}\ ,h_t$ through the recursion

$$\log h_r = \alpha + \delta\log h_{r-1} + \sigma_v v_r\ \ (r = u+1,\mathrm{K}\ \ t)$$

using simulated v_r. For each set of $\left\{h_r\right\}_{r=u+1}^t$ so drawn evaluate

$$(2\pi)^{-(t-u)/2}\prod_{s=u+1}^t h_s^{-3/2}\exp\left[-\sum_{s=u+1}^t\left(y_s-\mathbf{x}_s'\beta\right)^2/2h_s\right]. \qquad (22)$$

Then, average (22) over the R sets of drawn $h_{u+1},\mathrm{K}\ ,h_t$. (For the reported results, R=10. However, R=1 does very nearly as well.) Finally, the grand mean over the M replications of $\left(\theta^{(m)},\mathbf{h}^{(m)}\right)$ provides the desired numerical approximation to the predictive likelihood.

Computations were carried out using M=2000 iterations of the Gibbs sampling - data augmentation algorithm. There was very little serial correlation in the sampled values of the marginalized predictive likelihoods, and numerical standard errors were about 2% of the predictive likelihood for $t-u=20$. Computation time ranged from just over 2 minutes for the smallest sample, to about 7.5 minutes for the largest sample, on a Sun 10/51.

5.3 Priors for U.S./Canadian exchange data, 1975-1986

The GARCH and stochastic volatility models were compared for the time series $y_t = 100\cdot\log\left(x_t/x_{t-1}\right)$ where x_t is the closing value of the Canadian dollar on business day t. The only covariate in either model is a constant. This data set has also been studied

using the stochastic volatility model by Jacquier, Polson and Rossi (1994). The data set was supplied by the authors.

The 757 observations in 1975-1977 formed the initial sample in this experiment. These observations, together with the protoprior distribution of the parameters in each model, form the prior distributions in the computation of posterior odds ratios for the period beginning in January, 1978, and extending through the different months through November, 1986. Proper protoprior distributions were used in both models.

In the GARCH model, $\beta \sim \mathrm{N}(0, .1^2)$. Since returns are measured in percentage points, this is a diffuse protoprior distribution for mean return. The protoprior distribution of (γ, δ) is uniform on the unit simplex. To specify the protoprior distribution for a, we consider reasonable values for the unconditional mean $\alpha/(1-\gamma-\delta)$ of h_t and then assign a large variance. Given the behavior of exchange rate returns generally, $\mathrm{E}(h_t^{1/2}) = .1$ is reasonable. If $\gamma + \delta \approx .9$ then $\alpha \approx (.1)^2(1-.9)$ and $a = \log(\alpha) \approx -6.8$. To allow substantial variation in a, let $\underline{s}_a = \log(9) = 2.1972$. We shall refer to these specifications as the base prior, or prior 0, for the GARCH model.

The first row of Table 2 provides the marginalized likelihood for May, 1982, corresponding to this protoprior and the prior sample extending through the first business day of May, 1982. Three different values are given, corresponding to three different seeds of the random number generator and therefore three different draws of an initial value $\theta^{(1)}$ from the protoprior distribution. May, 1982, was chosen for study because this month is about midway through the 1978-1986 likelihood period: one would expect more sensitivity to the protoprior before this point, and less sensitivity after. In parentheses beside each marginalized likelihood evaluation is the corresponding numerical standard error, computed as described in Geweke (1992). Notice that the differences in the three evaluations of the marginalized likelihood are consistent with their numerical standard errors.

We assess the sensitivity of the marginalized likelihood to the protoprior by making changes in the protoprior distribution. In protoprior 1, $\beta \sim \mathrm{N}(0, .01^2)$ instead of $\beta \sim \mathrm{N}(0, .1^2)$; in protoprior 2, corresponding to an unconditional mean of .2 for $h_t^{1/2}$, $a \sim \mathrm{N}(-5.4, 2.1972^2)$ rather than $a \sim \mathrm{N}(-6.8, 2.1972^2)$; in protoprior 3, $\underline{s}_a = \log(4)$ so that $a \sim \mathrm{N}(-6.8, 1.3863^2)$; in protoprior 4, $\pi(\alpha, \delta) \propto |\alpha + \delta|$ rather than $\pi(\alpha, \delta) = 2$ on the unit simplex. In each case only one parameter of the protoprior distribution was changed, while other values remained at the protoprior 0 levels. Protoprior 5 makes all these changes simultaneously. We also examine sensitivity of the initialization specification of h_1. The base specification used $h_1 = \sum_{t=1}^{1030} (y_t - \bar{y})^2 / 1030$. In variant 6 we scale this value by 4, in variant 7 by .25, and in variant 8 we take $h_1 = 0$, all the time maintaining protoprior 0.

The results of these experiments are shown in the upper panel of Table 2. In every row, differences among the three evaluations are consistent with numerical standard errors. Comparing rows, it is evident that only prior 1 changes the evaluation of the marginalized likelihood in a detectable way. (The effect of the prior 2 change is exhibited in row 5 as well as row 2.) In the posterior distribution for the first business day of May, 1982, corresponding to prior 0, $\beta \sim \mathrm{N}(-.0114, .0046^2)$. Prior 1 is informative relative to this posterior distribution. As one would expect, the effect of prior 1 is to lower the marginalized likelihood (by about 6%).

In the stochastic volatility model the protoprior specifies $\beta \sim \mathrm{N}(0, .1^2)$ and $\delta \sim \mathrm{N}(.8, .3^2)$. The same unconditional mean and variance for $h_t^{1/2}$ then leads to a bivariate normal density for (α, δ) by means of (21c). The unconditional variance for h_t is $\sigma_v^2/(1-\delta^2)$ in the stochastic volatility model; evaluation at $\delta = .8$ then leads to $\underline{s}_v = .889$ in (21b), and we specify $\nu = 1$ to make the protoprior diffuse. Evaluations of marginalized likelihood for May, 1982, corresponding to this protoprior are presented in line 0 of the bottom panel of Table 2.

We consider eight variants of the protoprior distribution: in protoprior 1, $\beta \sim \mathrm{N}(0, .01^2)$; in protoprior 2, $\delta \sim \mathrm{N}(.95, .3^2)$ and in protoprior 3 $\delta \sim \mathrm{N}(.8, .15^2)$; in protoprior 4 the unconditional mean of $h_t^{1/2}$ is changed from .1 to .2, and in protoprior 5 the standard deviation of the unconditional mean of $\log(h_t)$ is changed from 3 to 2; in protoprior 6 $\underline{s}_v$ is changed from .889 to .561 consistent with the change in the unconditional mean of $h_t^{1/2}$ from .1 to .2; and in protoprior 7, $\underline{\nu}_v$ is increased from 1 to 10, thus making the original protoprior more informative. In each of these cases, only one parameter of the protoprior distribution was changed, while values of the other parameters remained at the prior 0 levels. Protoprior 8 makes all these changes simultaneously.

The results of these experiments are shown in the lower panel of Table 2. Differences among evaluations within rows are again consistent with numerical standard errors. Numerical standard errors are higher here than in the GARCH model, which makes differences between rows harder to detect. There is some evidence that the tighter distribution for β (prior 1) once again has some impact on the marginalized likelihood, but there is no indication of the changes produced by the other protoprior distributions.

5.4 Comparison using U.S./Canadian exchange data, 1975-1986

We computed marginalized likelihoods and Bayes factors in one-month increments. In each case the sample extends from Friday, January 3, 1975, through the first business day of the month indicated. In each model predictive likelihoods are then formed for the next day, for the next two days, and so on until all business days through the first business day

of the next month are incorporated. Thus the computational procedure just described was repeated for 106 samples, the smallest ending Tuesday, January 3, 1978, and the largest ending Monday, November 3, 1986. For each sample 20 to 23 functions of interest are computed, corresponding to predictive likelihoods ending on each business day in the month ahead.

The results of this procedure are displayed in Figure 1. The upper panel indicates the predictive likelihood for the data in the month indicated plus the first business day of the next month, based on the posterior distribution for the sample extending through the first working day of the month indicated. For example, using the posterior as of the end of the day Tuesday, January 3, 1978, the predictive odds ratio over observations from Wednesday, January 4, 1978 through Wednesday, February 1, 1978, is exp(2.05)=7.79. Using the posterior as of February 1, 1978, the predictive odds ratio over observations from Thursday, February 2, 1978, through Wednesday, March 1, 1978, is exp(-.484)=0.62. Hence the predictive odds ratio in favor of the stochastic volatility model for the first two months of 1978 is 7.77x0.62=4.80=exp(1.57). The lower panel of Figure 1 indicates the cumulative log Bayes factor in favor of the stochastic volatility model.

There is substantial variation in the monthly predictive Bayes factor. On the whole, however, evidence in favor of the stochastic volatility model steadily mounts with additional data. In every year except 1982, the marginal evidence for that year provides a Bayes factor greater than 1:1 in favor of this model. Over the period of nearly nine years, the predictive Bayes factor in favor of the stochastic volatility model is 1.89×10^{15}:1.

The decomposition of this Bayes factor into predictive factors, as described in Section 2, provides some insight into this result. Observe that for a few months, the Bayes factor in favor of the stochastic volatility model is quite large. Four months -- October, 1978 with a log Bayes factor of 5.93; February, 1979, 4.05; May, 1980, 7.43; and February, 1985, 5.67 -- account for nearly two-thirds of the final log Bayes factor of 35.17. Greater detail for each of these four months is displayed in Figures 2 through 5, respectively. Each figure consists of three panels. The upper panel indicates the predictive likelihood through each business day of the month for both models. The middle panel shows the corresponding Bayes factor in favor of the stochastic volatility model, and the bottom panel plots the return for each business day. For October, 1978 (Figure 2), the large Bayes factor is due to one day, Friday, October 27, when the exchange rate moved by almost 1%, or about five times a typical movement in the weeks preceding. For February, 1979 (Figure 3), more than half of the log predictive Bayes factor arises on Thursday, March 1, the last day of the predictive horizon. Once again, there was a very large movement in exchange rates on that day. Figures 4 and 5 tell similar stories. The May 1980 contribution of 7.43 was made almost

entirely on Wednesday, May 21 when exchange rates moved by more than 1%. And in February, 1985, essentially the entirely monthly contribution to the Bayes factor was realized on Thursday, February 21, when the rate fell by over 1%. About the same change occurred on the next two business days. Both subsequent events were unlikely under either model (note the declines in the marginalized predictive likelihoods on each day) but not nearly so unlikely as the first of the three large shocks in the context of the GARCH model.

Figures 6 and 7 display similar detail for the two months in which the GARCH model fared best relative to the stochastic volatility model, but the pattern here is different. There is no single day or even a few days that account for the relatively better performance of GARCH. Returns are quite volatile throughout both months, indicating that h_t is probably large in both models. The large movement on June 24 is improbable in both models, but only slightly moreso in the stochastic volatility model.

Finally, Figure 8 provides detail for a month in which the predictive Bayes factor was about 1.0. This month is relatively tranquil. Throughout the month h_t declines in each model, and since tranquillity prevails all marginalized likelihoods in both models rise slowly. Since this happens at the same rate, the Bayes factor is not much affected.

This analysis of the decomposition of the Bayes factor suggests three features that are key in accounting for the relative performance of the models. First, both models perform about as well in periods of sustained tranquillity (Figure 8). Second, GARCH may perform slightly better in periods of sustained volatility (Figures 6 and 7), but this is at most a tentative conjecture on the basis of this analysis. Third, and by far the most important, the observed large and sudden movements in exchange rates are relatively much more probable in the stochastic volatility model than in GARCH. Our casual inspection, beginning with Figure 1, uncovered four one-day events that were, respectively, 525, 8.7, 584, and 232 times more probable in the stochastic volatility model than in the GARCH model. These events are highly informative.

It would appear that the stochastic volatility model accommodates larger movements better for two reasons. Most important, since conditional variance in the stochastic volatility model is a stochastic function of past conditional variances, sudden relatively large shocks are plausible if they are not too large relative to the standard deviation (σ_v) of the innovation of the conditional variance process. Since conditional variance is a deterministic function of the past in the GARCH model, sudden relatively large shocks must appear implausible. A second, contributing reason is the logarithmic form of the evolution of conditional variance in the stochastic volatility model, which makes new, larger conditional variances more plausible than if the innovation to the conditional variance process were additive. Of course, not all models in the ARCH family maintain an additive functional form for the evolution of

conditional variance from its own past and past returns, but so long as this evolution is deterministic they are likely to be subject to the limitations uncovered here. A promising direction for further development in both models might be the introduction of leptokurtic shocks.

6. An example: Limited information Bayesian inference

We now take up a second illustration of computation of marginalized likelihoods and Bayes factors. The objectives are to see how practical the procedures are in a model with many parameters in a very small sample. One expects computational problems to be greater in this situation than in models with few parameters and many observations, which was the case in the example in Section 5. This provides a good opportunity to measure the computational advantages inherent in the decomposition of the marginalized likelihood into a product of predictive Bayes factors.

6.1 The model

The linear simultaneous equation model may be written

$$\underset{t\times g}{\tilde{Y}}\ \underset{g\times g}{\Gamma} = \underset{t\times m}{\tilde{X}}\ \underset{m\times g}{\mathrm{B}} + \underset{t\times g}{H}. \tag{23}$$

Each row denotes an observation and each column an equation. The model as a whole determines the g endogenous variables in the columns of $\tilde{Y}$, given the m predetermined variables in the columns of $\tilde{X}$ and disturbances in the columns of H:

$$\tilde{Y} = \tilde{X}\mathrm{B}\Gamma^{-1} + H\Gamma^{-1} = \tilde{X}R + H^{*}. \tag{24}$$

In the language of the simultaneous equation literature (23) is the structural form of the simultaneous equation system. Each equation corresponds to an aspect of economic behavior (for example, a supply or demand equation), and typically does not contain all -- or even most -- of the variables in the system. This implies restrictions to zero of corresponding elements of Γ and B. If these restrictions take the proper form then one may determine Γ and B from knowledge of R in the multivariate regression (24). The simultaneous equation literature provides extensive treatment of these identification conditions.

The example here concentrates on one of the g equations, without loss of generality the first:

$$\underset{t\times L}{Y}\ \underset{L\times 1}{\gamma} = \underset{t\times k}{X}\ \underset{k\times 1}{\beta} + \underset{t\times 1}{\varepsilon}. \tag{25}$$

Here, Y includes exactly those $L\,(\leq g)$ endogenous variables that appear in the first equation -- that is, those for which the corresponding elements of the first column of Γ are

not restricted to be zero. The vector γ contains these nonzero elements. Similarly X consists of those $k\,(\le m)$ predetermined variables in the first equation, and the corresponding nonzero coefficients from the first column of B appear in β. Without loss of generality suppose that these are the predetermined variables in the first k columns of $\tilde{X}$ in (23). In general, it is the case that $k^* = m - k \ge L - 1$: this is a necessary condition for identification of the elements of γ and β from R in (24) if exclusions are the only type of restriction in the system (Theil, 1971, pp. 448-49). The coefficients in γ and β must be normalized, and this is usually done by taking $\gamma_1 = 1$.

Now consider the L equations in the reduced form (24) corresponding to the L endogenous variables included in the first structural equation (25). Write this subsystem

$$Y = \underset{n\times k}{X} \quad \underset{n\times k^*}{X^*} \begin{bmatrix} \underset{k\times L}{\Pi} \\ \underset{k^*\times L}{\Pi^*} \end{bmatrix} + E, \quad \mathrm{vec}(E) \sim \mathrm{N}\left(0, \Sigma \otimes \mathrm{I}_n\right). \tag{26}$$

The matrix X^* consists of the last $k^* = m - k$ columns of X, corresponding to the predetermined variables not included in the first equation. From (25),

$$\underset{k\times L}{\Pi}\ \underset{L\times 1}{\gamma} = \underset{k\times 1}{\beta} \text{ and } \underset{k^*\times L}{\Pi^*}\ \underset{L\times 1}{\gamma} = \underset{k^*\times 1}{\mathbf{0}}.$$

For the second of these equations to be satisfied it must be the case that $\mathrm{Rank}\left(\Pi^*\right) \le L - 1$. When $k^* = L - 1$ this is trivially true, but if $k^* > L - 1$ (which tends to be the rule) there is a rank restriction on Π^*. With this in mind rewrite (26) as

$$Y = \begin{bmatrix} X & X^* \end{bmatrix} \begin{bmatrix} \Pi \\ \Psi\Phi \end{bmatrix} + E = \underset{n\times k}{X}\ \underset{k\times L}{\Pi} + \underset{n\times xk^*}{X^*}\ \underset{k^*\times(L-1)}{\Psi}\ \underset{(L-1)\times L}{\Phi} + \underset{n\times L}{E} \tag{27}$$

The system (27) is a modest extension of the reduced rank regression model, for which Bayesian inference using a Gibbs sampling algorithm to compute posterior moments is taken up in Geweke (1993). Given suitable normalization of Ψ or Φ and suitable independent prior densities for Π, Ψ, Φ and Σ, the conditional posterior densities for Π, Ψ and Φ are all multivariate normal and that for Σ is inverted Wishart. Note that conditional on Ψ, Φ and Σ,

$$Y - X^*\Psi\Phi = X\Pi + E$$

is a multivariate regression model, implying a simple conditional posterior distribution for Π (Zellner, 1971, p. 227). Conditional distributions for the elements of Φ and Ψ are respectively multivariate normal, but the distributions are more complicated (Geweke, 1993).

The functions of interest generally will be

$$\gamma : \Psi\Phi\gamma = 0, \quad \beta = \Pi\gamma, \text{ and } \sigma^2 = \mathrm{var}(\varepsilon) = \gamma\Sigma\gamma.$$

The first equation may be solved through a singular value decomposition of $\Psi\Phi$, followed by normalization of γ, and then β and σ^2 follow directly. For the purposes of illustration

in this paper, we are interested in comparing (27) with the unrestricted multivariate regression model,

$$Y = X\mathrm{A} + U, \quad \mathrm{vec}(U) \sim \mathrm{N}(0, \Omega \otimes \mathrm{I}_n). \tag{28}$$

For (27) we employ the improper protoprior kernel $\mathrm{f}_{1s_0}(\Pi, \Psi, \Phi, \Sigma) \propto \left|\Sigma^{-1/2}\right|$, and for (28) the improper protoprior kernel $\mathrm{f}_{2s_0}(\mathrm{A}, \Omega) \propto \left|\Omega^{-1/2}\right|$. Marginalized likelihoods are computed in two ways: by numerical approximation of each of the predictive likelihoods in (8) followed by computation of the product, which we shall call the "linked" method; and by direct numerical approximation of (3), which we shall call the "raw" method. The considerations raised in Section 3 suggest that the linked method should be more accurate than the raw method.

6.2 Demand for meat (Tintner)

A classic example found in the simultaneous equation literature is Tintner's (1965, p. 176) meat market model, which consists of a demand equation and a supply equation:

$$\gamma_{11} y_1 + \gamma_{21} y_2 = \beta_{11} + \beta_{21} x_2 + u_1 \tag{29a}$$

$$\gamma_{12} y_1 + \gamma_{22} y_2 = \beta_{12} + \beta_{32} x_3 + \beta_{42} x_4 + u_2 \tag{29b}$$

where y_1 is the quantity of meat consumed, y_2 is the price of meat, x_2 is per capita disposable income, x_3 is the cost of processing meat, and x_4 is the cost of producing agricultural products. Annual data (23 observations) for the period 1919-41 are used. A summary of the data is found in Tintner (1965, pp. 177-78). This work begins with the actual observations given in French (1950, p. 27).

In the case of the demand equation $L = k = k^* = 2$, and therefore the matrix Π^* must have rank 1. In the case of the supply equation, $L = 2, k = 3, k^* = 1$. The 1×2 matrix Π^* therefore has rank 1 without further restrictions. (In the language of the simultaneous equation literature, the first equation is "overidentified" and the second is "just identified.") Consequently the joint posterior distribution for all the parameters in the system (29) follows from the posterior distribution of Π and Π^* as constructed from the demand equation (29a) alone.

Tables 3 and 4 present marginalized likelihoods, and Table 5 presents Bayes factors, for a variety of choices of s_0 (the last observation entering the prior) and t (the last observation entering the likelihood) and for alternative computational procedures. The Bayes factors are expressed in favor of the restricted model (27). The unrestricted model has 11 parameters, and the smallest sample forming a prior (as of 1934) has 16 observations. Computations employed 10^4 iterations of an independence Monte Carlo sampler for the multivariate regression model, requiring about 12 seconds on a Sun 10/51.

For the reduced rank regression model (27), 10^5 iterations were used with functions of interest evaluated every 10th iteration. This required about 4 minutes.

Tables 3 and 4 show marginalized likelihoods for the reduced rank model (27) and the multivariate regression model (28) respectively. In each table raw marginalized likelihoods are presented first, followed by linked marginalized likelihoods. Observe that marginalized likelihoods in the two models tend to move together: the relative predictive probability of observations and groups of observations is less volatile than are the prediction probabilities themselves. Without exception, numerical standard errors for the linked marginalized likelihoods are lower than the numerical standard errors for the corresponding raw marginalized likelihoods. Relative computed accuracy varies substantially from observation to observation, but on the whole the relative advantage of the linked procedure is greater than it is for the normal case documented in Table 1. Linked and raw results agree up to computed numerical standard errors, and in all cases the linked marginalized likelihoods are accurate to at least one significant figure.

Bayes factors in favor of the restricted model (27) are displayed in Table 5. In virtually all cases numerical standard error is smaller relative to the corresponding Bayes factor for the linked method than it is for the raw method. However in many cases linked Bayes factors are about double the corresponding Bayes factors, resulting in somewhat higher numerical standard errors for the linked Bayes factors. Precisely as one would expect from the analysis of Section 4, the relative accuracy of Bayes factors diminishes as the likelihood function involves more products . Numerical standard error ranges from about 2% of the Bayes factor (for 1941, based on the 1919-40 sample) to as much as 15% of the Bayes factor (for 1935-41, based on the 1919-34 sample).

7. Conclusion

This paper has developed a general method for the computation of the marginalized likelihood and demonstrated its practicality in some models typical of those used in macroeconomics and finance. From reported marginalized likelihoods it is easy to form posterior odds ratios between competing models, thus enabling their comparison. It is no more difficult to compare several models than it is to compare two. It is irrelevant whether models are nested or not, and the example in Section 5 shows that they can be quite different. All that is required is that the same data set be used. The method proceeds by decomposing marginalized likelihood, and therefore posterior odds ratios, into period-by-period components. As documented in one of the examples, analysis of this decomposition can provide an interpretation of marginalized likelihoods and posterior odds ratios that may

be fruitful in understanding shortcomings in the models considered, and in suggesting promising lines of further development.

This procedure builds on simulation methods for computing posterior moments whose recent rapid development has greatly expanded the class of models that may be used in formal Bayesian inference. These simulation methods strongly suggest a style of public reporting with several practical attractions. In particular, making available the simulated parameter values enables any other investigator to employ different priors, update the original posterior with new data, evaluate the posterior expectations of new functions of interest, or any combination of these three, by means of simple computations that are not model specific.

These procedures may enable economists and decision makers to bring the full power of the Bayesian paradigm to bear on a wide array of problems, providing results more directly relevant to these problems than is now the case. (For example, it is a short step from the results reported in Section 5 to the valuation of options, futures, and options contracts on futures.) They may even reinvigorate the link between econometrics and practical decision making that was emphasized at the inception of our profession. That would be a welcome development, indeed.

References

Atkinson, A.C., 1978, "Posterior Probabilities for Choosing a Regression Model," *Biometrika* **65**: 39-48.

Bartlett, M.S., 1957, "A Comment on D.V. Lindley's Statistical Paradox," *Biometrika* **44**: 533-534.

Berger, J.O., and L.R. Pericchi, 1992, "The Intrinsic Bayes Factor," Purdue University Department of Statistics technical report.

Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* **31**: 307-327.

Bollerslev, T., R. Chou, and K.F. Kroner, 1992, "ARCH Modeling in Finance," *Journal of Econometrics* **52**: 5-59.

Casella, G. and E.I. George, 1992, "Explaining the Gibbs Sampler," *The American Statistician* **46**: 167-174.

Chib, S. and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," Washington University working paper.

DeGroot, M., 1970, *Optimal Statistical Decisions*. New York: McGraw-Hill.

Engle, R., 1982, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* **50**: 987-100.

French, B. L., 1950, "Application of Simultaneous Equations to the Analysis of the Demand for Meat." Unpublished M.A. thesis, Iowa State University.

Gelfand, A.E., and D.K. Dey, 1994, "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society Series B* **56**: 501-514.

Gelfand, A.E., and A.F.M. Smith, 1990, "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* **85**: 398-409.

Gelman, A., and D.B. Rubin, 1992, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science* **7**: 457-472.

Geman, S., and D. Geman, 1984, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721-741.

Geweke, J., 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* **57**: 1317-1340.

Geweke, J., 1992, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J.M. Bernardo *et al.* (eds.), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*. Oxford: Clarendon Press.

Geweke, J., 1993, "Bayesian Reduced Rank Regression in Econometrics," Federal Reserve Bank of Minneapolis working paper.

Geweke, J., 1994, "Monte Carlo Simulation and Numerical Integration," chapter forthcoming in H. Amman, D. Kendrick and J. Rust (eds.), *Handbook of Computational Economics*. Federal Reserve Bank of Minneapolis Research Department working paper 526, April.

Geweke, J., M. Keane, and D. Runkle, 1994, "Alternative Computational Approaches to Statistical Inference in the Multinomial Probit Model," Federal Reserve Bank of Minneapolis Staff Report 170, May.

Geyer, C.J., 1992, "Practical Markov Chain Monte Carlo," *Statistical Science* **7**: 473-481.

Hannan, E.J., 1970, *Multiple Time Series*. New York: Wiley.

Hastings, W.K., 1970, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika* **57**: 97-109.

Hildreth, C., 1963, "Bayesian Statisticians and Remote Clients," *Econometrica* **31**: 422-438.

Jacquier, E., N.G. Polson, and P.E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, forthcoming.

Jeffreys, H., 1961, *Theory of Probability*. Oxford: Clarendon. (Third edition)

Lindley, D.V., 1957, "A Statistical Paradox," *Biometrika* **44**: 187-192.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, 1954, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics* **21**: 1087-1092.

Newton, M.A., and A.E. Raftery, 1994, "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society Series B*, forthcoming. (Also University of Washington Department of Statistics technical report.)

O'Hagan, A., 1991, "Discussion on Posterior Bayes Factors (by M. Aitkin)", *Journal of the Royal Statistical Society Series B*, **53**: 136.

Roberts, G.O., and A.F.M. Smith, 1992, "Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms," University of Cambridge Statistical Laboratory Research Report No. 92-30.

Spiegelhalter, D.J. and A.F.M. Smith, 1982, "Bayes Factors for Linear and Log-linear Models with Vague Prior Information," *Journal of the Royal Statistical Society Series B*, **44**: 377-387.

Tanner, M.A., and W.H. Wong, 1987, "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association* **82**: 528-550.

Taylor, S., 1986, *Modeling Financial Time Series*. New York: John Wiley and Sons.

Theil, H., 1971, *Principles of Econometrics.* New York: Wiley.

Tierney, L., 1991, "Markov Chains for Exploring Posterior Distributions," Technical Report No. 560, University of Minnesota School of Statistics. Forthcoming, *Annals of Statistics*.

Tintner, G., 1965, *Econometrics* (Second edition). New York: Wiley.

Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Table 1

[Expected numerical standard error]/[Expected numerical standard error with linking]
as a function of sample size (t) and observations in likelihood function (r)
with k=9 parameters

	t=9	t=20	t=50
r=2	1.186	1.079	1.031
r=3	1.418	1.166	1.062
r=5	2.076	1.369	1.130
r=10	6.082	2.108	1.326
r=20	7.487	5.609	1.863
r=50	3.953×10^5	1.913×10^2	6.045

Table 2

Sensitivity of marginalized likelihood for May, 1982,
to prior specification, initial conditions, and likelihood initialization[a]

GARCH

Initial conditions Prior/ initialization	a	b	c
0	.0358316 (.0001987)	.0359298 (.0002116)	.0355990 (.0001980)
1	.0340285 (.0002047)	.0340550 (.0002096)	.0334190 (.0001854)
2	.0357066 (.0002090)	.0354063 (.0001850)	.0354527 (.0001927)
3	.0357828 (.0001893)	.0360759 (.0002130)	.0356769 (.0001587)
4	.0358046 (.0001641)	.0360037 (.0001868)	.0354254 (.0002301)
5	.0334715 (.0001601)	.0335545 (.0001601)	.0333079 (.0001979)
6	.0359182 (.0002309)	.0355570 (.0002299)	.0358072 (.0001798)
7	.0355307 (.0001603)	.0351651 (.0001927)	.0352200 (.0001205)
8	.0356016 (.0001639)	.0355034 (.0001669)	.0355820 (.0001541)

Stochastic volatility

Initial conditions Prior/ initialization	a	b	c
0	.0294394 (.0007781)	.0298064 (.0007450)	.0300387 (.0006926)
1	.0281328 (.0007101)	.0287615 (.0007187)	.0270598 (.0005363)
2	.0286919 (.0006656)	.0300395 (.0007169)	.0286605 (.0007484)
3	.0305217 (.0006101)	.0299146 (.0006883)	.0310686 (.0007175)
4	.0293207 (.0007710)	.0280117 (.0007414)	.0303610 (.0006978)
5	.0289205 (.0006436)	.0296276 (.0005878)	.0291596 (.0005850)
6	.0295256 (.0007364)	.0297951 (.0005424)	.0287858 (.0007298)
7	.0317184 (.0007401)	.0305197 (.0006872)	.0295160 (.0005276)
8	.0297489 (.0006157)	.0315650 (.0007040)	.0301338 (.0006945)

[a]Prior specifications and likelihood initializations are described in the text. Alternative initial conditions entail different seeds for the random number generator, and a different draw of the initial parameter vector from the prior distribution. Numerical standard errors of the numerical approximations to the marginalized likelihoods are indicated parenthetically.

Table 3

Marginalized likelihoods for Meat model, Multivariate regression

Raw computations

	t=1935	t=1936	t=1937	t=1938	t=1939	t=1940	t=1950
s_0=1934	$4.603\text{x}10^{-3}$ (.088)	$2.006\text{x}10^{-4}$ (.038)	$3.633\text{x}10^{-6}$ (.106)	$6.111\text{x}10^{-11}$ (.775)	$2.171\text{x}10^{-12}$ (.289)	$3.299\text{x}10^{-14}$ (.653)	$3.342\text{x}10^{-16}$ (1.025)
s_0=1935		$4.330\text{x}10^{-2}$ (.015)	$7.724\text{x}10^{-4}$ (.085)	$1.568\text{x}10^{-8}$ (.179)	$5.522\text{x}10^{-10}$ (.660)	$7.032\text{x}10^{-12}$ (.496)	$6.551\text{x}10^{-14}$ (.623)
s_0=1936			$1.765\text{x}10^{-2}$ (.022)	$3.109\text{x}10^{-7}$ (.222)	$1.081\text{x}10^{-8}$ (.105)	$1.506\text{x}10^{-10}$ (.200)	$1.494\text{x}10^{-12}$ (.293)
s_0=1937				$1.721\text{x}10^{-5}$ (.120)	$6.168\text{x}10^{-7}$ (.388)	$8.838\text{x}10^{-9}$ (.597)	$8.141\text{x}10^{-11}$ (.759)
s_0=1938					$3.514\text{x}10^{-2}$ (.011)	$4.892\text{x}10^{-4}$ (.036)	$4.967\text{x}10^{-6}$ (.124)
s_0=1939						$1.396\text{x}10^{-2}$ (.007)	$1.423\text{x}10^{-4}$ (.028)
s_0=1940							$1.012\text{x}10^{-2}$ (.010)

Linked predictive factors

	t=1935	t=1936	t=1937	t=1938	t=1939	t=1940	t=1950
s_0=1934	$4.603\text{x}10^{-3}$ (.088)	$1.993\text{x}10^{-4}$ (.039)	$3.518\text{x}10^{-6}$ (.081)	$6.054\text{x}10^{-11}$ (.445)	$2.127\text{x}10^{-12}$ (.156)	$2.970\text{x}10^{-14}$ (.219)	$3.009\text{x}10^{-16}$ (.224)
s_0=1935		$4.330\text{x}10^{-2}$ (.015)	$7.642\text{x}10^{-4}$ (.099)	$1.315\text{x}10^{-8}$ (.093)	$4.622\text{x}10^{-10}$ (.328)	$6.452\text{x}10^{-12}$ (.459)	$6.536\text{x}10^{-14}$ (.470)
s_0=1936			$1.765\text{x}10^{-2}$ (.022)	$3.038\text{x}10^{-7}$ (.215)	$1.067\text{x}10^{-8}$ (.076)	$1.490\text{x}10^{-10}$ (.106)	$1.509\text{x}10^{-12}$ (.108)
s_0=1937				$1.721\text{x}10^{-5}$ (.120)	$6.048\text{c}10^{-7}$ (.422)	$8.442\text{x}10^{-9}$ (.591)	$8.552\text{x}10^{-11}$ (.604)
s_0=1938					$3.514\text{x}10^{-2}$ (.011)	$4.906\text{x}10^{-4}$ (.029)	$4.969\text{x}10^{-6}$ (.057)
s_0=1939						$1.396\text{x}10^{-2}$ (.007)	$1.414\text{x}10^{-4}$ (.016)
s_0=1940							$1.013\text{x}10^{-2}$ (.010)

The mantissa of the corresponding numerical standard error is indicated in parentheses directly below the numerical approximation of the marginalized likelihood. The exponent of the numerical standard error is the same as that of the approximated marginalized likelihood.

Table 4

Marginalized likelihoods for Meat model, Reduced rank regression

Raw computations

	t=1935	t=1936	t=1937	t=1938	t=1939	t=1940	t=1950
s_0=1934	1.257×10^{-3} (.178)	4.934×10^{-5} (.781)	1.038×10^{-6} (.219)	3.541×10^{-11} (.830)	1.309×10^{-12} (.309)	2.103×10^{-14} (.526)	2.318×10^{-16} (.650)
s_0=1935		3.852×10^{-2} (.082)	7.166×10^{-4} (.421)	3.328×10^{-8} (.355)	1.186×10^{-9} (.129)	1.698×10^{-11} (.177)	1.823×10^{-12} (.212)
s_0=1936			2.005×10^{-2} (.069)	7.818×10^{-7} (.041)	2.643×10^{-8} (.205)	3.868×10^{-10} (.261)	4.591×10^{-12} (.474)
s_0=1937				3.632×10^{-5} (.262)	1.280×10^{-6} (.089)	1.849×10^{-8} (.139)	1.924×10^{-10} (.175)
s_0=1938					3.532×10^{-2} (.015)	4.959×10^{-4} (.044)	5.017×10^{-6} (.088)
s_0=1939						1.404×10^{-2} (.011)	1.473×10^{-4} (.025)
s_0=1940							1.055×10^{-2} (.015)

Linked predictive factors

	t=1935	t=1936	t=1937	t=1938	t=1939	t=1940	t=1950
s_0=1934	1.257×10^{-3} (.178)	4.842×10^{-5} (.693)	9.708×10^{-7} (1.43)	3.526×10^{-11} (.578)	1.245×10^{-12} (.204)	1.749×1^{-14} (.287)	1.845×10^{-16} (.304)
s_0=1935		3.835×10^{-2} (.082)	7.723×10^{-4} (.313)	2.805×10^{-8} (.232)	9.908×10^{-10} (.821)	1.391×10^{-11} (.116)	1.468×10^{-13} (.124)
s_0=1936			2.005×10^{-2} (.069)	7.828×10^{-7} (.582)	2.572×10^{-8} (.206)	3.611×10^{-10} (.290)	3.810×10^{-12} (.311)
s_0=1937				3.623×10^{-5} (.262)	1.283×10^{-6} (.0930	1.801×10^{-8} (.131)	1.900×10^{-10} (.141)
s_0=1938					3.532×10^{-2} (.015)	4.959×10^{-4} (.044)	5.232×10^{-6} (.088)
s_0=1939						1.404×10^{-2} (.011)	1.481×10^{-4} (.024)
s_0=1940							1.055×10^{-2} (.015)

The mantissa of the corresponding numerical standard error is indicated in parentheses directly below the numerical approximation of the marginalized likelihood. The exponent of the numerical standard error is the same as that of the approximated marginalized likelihood.

Table 5

Bayes factors for Meat model

Raw computations

	t=1935	t=1936	t=1937	t=1938	t=1939	t=1940	t=1950
s_0=1934	.2731 (.0390)	.2464 (.0392)	.3600 (.0612)	.5795 (.1545)	.6029 (.1634)	.6390 (.2040)	.6936 (.2882)
s_0=1935		.8896 ((.0192)	.9278 (.0554)	2.122 (.3312)	2.148 (.347)	2.415 (.304)	2.783 (.418)
s_0=1936			1.136 (.042)	2.515 (.220)	2.445 (.304)	2.568 (.383)	3.073 (.681)
s_0=1937				2.110 (.212)	2.075 (.195)	2.092 (.157)	2.363 (.308)
s_0=1938					1.003 (.005)	1.014 (.0117)	1.010 (.031)
s_0=1939						1.006 (.009)	1.035 (.027)
s_0=1940							1.041 (.018)

Linked predictive factors

	t=1935	t=1936	t=1937	t=1938	t=1939	t=1940	t=1950
s_0=1934	.2731 (.0390)	.2430 (.0351)	.2760 (.0411)	.5824 (.1047)	.5853 (.1053)	.5889 (.1060)	.6132 (.1109)
s_0=1935		.8896 ((.0192)	1.011 (.043)	2.133 (.232)	2.144 (.234)	2.156 (.236)	2.246 (.249)
s_0=1936			1.136 (.042)	2.397 (.256)	2.410 (.258)	2.423 (.260)	2.525 (.275)
s_0=1937				2.110 (.212)	2.121 (.213)	2.133 (.215)	2.222 (.227)
s_0=1938					1.003 (.005)	1.011 (.011)	1.053 (.021)
s_0=1939						1.006 (.009)	1.047 (.021)
s_0=1940							1.041 (.018)

The mantissa of the corresponding numerical standard error is indicated in parentheses directly below the numerical approximation of the Bayes factor. The exponent of the numerical standard error is the same as that of the approximated Bayes factor.