Bayesian Analysis (2006) **1**, Number 3, pp. 547–550

Rejoinder

William J. Browne*, and David Draper†

We are grateful to Gelman, Kass and Natarajan, and Lambert for their thoughtful comments (and indeed for the original research that they summarize in their papers), and we offer the following remarks by way of rejoinder.

- Many of the results presented in our article were obtained more than a few years ago (based, as they were, on part of the work in Browne (1998)) and are only now seeing the light of publication largely due to, shall we say, the vagaries of non-Bayesian refereeing. We focused on the $\Gamma^{-1}(\epsilon, \epsilon)$ prior for random-effects variances in some of our work because—under the influence of the `WinBUGS` package and the examples distributed with it—this was very much the most common prior in use in hierarchical/multilevel modeling in the mid to late 1990s. Lambert expresses the opinion that this is still true today, although it appears to us that the pendulum is shifting away from this prior, for reasons like those mentioned by Gelman. (To be fair to the `WinBUGS` development group, in many of the examples distributed with release 1.4.1 they currently offer analyses with both $\Gamma(0.001, 0.001)$ priors on random-effects precision parameters τ and Uniform priors on the corresponding standard deviation parameters $\sigma = \tau^{-1/2}$, although they send a distinctly mixed message by building in default values of 0.001 for each of the shape and scale parameters whenever a parameter is given a Gamma distribution in the `DoodleBUGS` part of the package.)

  It is interesting to see that in 2006 there is still no consensus on a general-purpose choice of diffuse prior for this situation, although the work summarized in both the Gelman and Kass-Natarajan contributions to this discussion may go some distance toward achieving this goal. We have found ourselves recently gravitating toward Uniform priors on random-effects standard deviations, which accord with one of Gelman's suggestions, although instead of using Uniform$(0, \infty)$ (or Uniform$(0, A)$ for huge A) we prefer Uniform$(0, c)$ where c is chosen just large enough not to truncate the marginal likelihood for σ (and, in an interesting resurrection of the sometimes appropriately maligned Gamma prior, c can often be chosen well by making a preliminary fitting with a $\Gamma^{-1}(0.001, 0.001)$ prior on σ^2 and looking at the marginal posterior for σ). It is also interesting that $\Gamma^{-1}(\epsilon, \epsilon)$ priors were originally chosen for computational convenience (through their conditional conjugacy), and the half t family mentioned by Gelman again has surfaced due to computational benefits, this time arising from model expansion. One of us (Browne (2004)) has also seen these benefits in a more complex random effects model, reinforcing Gelman's comments on efficiency of MCMC chains.

*Division of Statistics, School of Mathematical Sciences, University of Nottingham, UK http://www.maths.nottingham.ac.uk/personal/pmzwjb/

†Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, http://www.ams.ucsc.edu/~draper

 ba0003

- The IGLS estimation method we use to get maximum likelihood estimates in the paper (see Section 2.1) has other features that are of interest to explore for their potential payoff with MCMC methods. Given a particular random effects model, the IGLS method does not in fact directly fit this model, but rather fits a structured multivariate normal model with the whole set of responses treated as one vector-valued outcome, and with constraints (e.g., positive between-groups variance) included in the covariance matrix of the response; these constraints create equivalence between the multivariate model and the original random effects model. We are currently investigating MCMC algorithms for such structured multivariate normal models; here we have the option of allowing the parameter in this model that corresponds to the between-groups variance in the random effects model to have positive prior probability of taking negative values. This has advantages in performing Bayesian model selection and may help in choosing a reference prior for this family of structured multivariate normal models (although the equivalence with the random effects model is lost by such a prior choice).

- In three places in Gelman's paper (Sections 5.1, 5.2, and 6.2) he refers to what he characterizes as the good performance of a particular choice of prior ("the simplest approach ... seems to perform well"; "this prior distribution appears to perform well in this example"; "the half-Cauchy prior distribution does slightly better than the uniform") without saying what standard of merit he is using to come to these conclusions. We believe that the best way to settle issues of this type is through simulation studies (of the type illustrated in our paper, in Kass and Natarajan's contribution to this discussion, and in Lambert et al. (2005)), in which an environment embodying a particular known truth is created and then a variety of Bayesian inferential methods are compared on their ability to reproduce the known truth. This is a form *calibration* inquiry—how often does my method get the right answer?—that it would seem all statisticians, whether they are using Bayesian methods or not, would be interested in undertaking. (How exactly can Gelman know that the half-Cauchy prior distribution does slightly better than the uniform in his ANOVA example without performing such a simulation? See, e.g., Draper (2006) for some recent thoughts on the importance of combining the notions of coherence (internal consistency) and calibration (external consistency) in contemporary Bayesian inference.) In fact, this simulation approach has by now become so easy to perform—e.g., by embedding calls to `WinBUGS` in a random-data-set-generating environment in `R` (in part thanks to the useful `R` functions Gelman has made available at `www.stat.columbia.edu/∼gelman/bugsR`)—and inexpensive computers have become so fast that most questions one might have about the calibration properties of a particular choice of diffuse prior can be answered in a completely problem-specific manner with just an hour or two of programming and a few hours or days of computer time.

At about the time of Browne (1998), we were the co-developers of the MCMC capabilities in the multilevel modeling package `MLwiN` (Rasbash et al. (2005)), and—since we wanted to give users a default choice of diffuse priors for that package—it was natural to ask calibration questions of the type addressed in our paper. We believe that similar questions are routinely worth asking, not just by software developers but by essentially all Bayesian analysts, and we hope that the implementation and publication of Bayesian calibration studies of the type discussed here will become considerably more frequent in the not-too-distant future.

References

Browne, W. J. 1998. *Applying MCMC Methods to Multilevel Models*. Ph.D. dissertation: University of Bath, U.K. 547, 548

—. 2004. An illustration of the use of reparameterisation methods for improving MCMC efficiency in crossed random effect models. *Multilevel Modelling Newsletter* **16**: 13–25. 547

Draper, D. 2006. Coherence and calibration: comments on subjectivity and "objectivity" in Bayesian analysis. Discussion of "The case for objective Bayesian analysis" by J. Berger and "Subjective Bayesian analysis: principles and practice" by M. Goldstein. *Bayesian Analysis* (this issue). 548

Lambert, P. C., A. J. Sutton, P. R. Burton, K. R. Abrams, and D. R. Jones. 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* **24**: 2401–2428. 548

Rasbash, J., F. Steele, W. Browne, and B. Prosser. 2005. *A User's Guide to MLwiN Version 2.0*. University of Bristol, U.K. (www.mlwin.com). 548