Eliciting a Prior Distribution for the Error Variance in Normal Linear Models

Fadlalla Elfadaly and Paul H. Garthwaite

Department of Mathematics and Statistics, The Open University,

Walton Hall, Milton Keynes, MK7 6AA, UK.

Summary

To elicit an informative prior distribution for a normal linear model, expert opinion must be quantified about both the regression coefficients and the random error. The latter task has attracted comparatively little attention. In this paper we develop a method of assessing a conjugate prior distribution for the error variance. The method quantifies an expert's opinions through assessments of a median and conditional medians. Interactive graphics is the medium through which assessments are elicited. A computer program that implements the method is available.

Keywords: Assessment tasks; Elicitation; Normal linear model; Prior distribution; Subjective probability.

1. Introduction

Methods for quantifying expert opinion about a normal linear model have been proposed by Kadane et al. (1980) and Garthwaite and Dickey (1988, 1992), amongst others. A normal linear model is, of course, a form of generalized linear model (GLM) and the task of quantifying opinion about GLMs has also been addressed (Bedrick, Christenson and Johnson, 1996; Chen et al., 2003; Garthwaite et al., 2011). These methods all focus primarily on the task of quantifying opinion about regression coefficients. For some GLMs, such as logistic regression, this determines a complete prior distribution but, with a normal linear model, prior opinion about the error variance must also be quantified in order to obtain a prior distribution for all model parameters.

Some of the above elicitation procedures contain a method of assessing opinion about a linear model's error variance, but the methods have drawbacks. For example, the method of Kadane et al. (1980) requires assessments of 0.9375 quantiles of predictive distributions in the part of their procedure that relates to the error variance. Assessing quantiles well into the tails of a distribution is a difficult task that people perform poorly (Alpert and Raiffa, 1969). If a different method were used to quantify opinion about the error variance, then that part of the procedure of Kadane et al. (1980) could be dropped. Garthwaite and Dickey (1988) separate the task of quantifying opinion about regression coefficients from that of quantifying opinion about the error variance. This is potentially beneficial; decomposing a complex assessment problem into a number of smaller problems is recommended (Hogarth, 1975). However the number of assessment Garthwaite and Dickey elicit from the expert is the minimum number needed to determine the hyperparameters of the prior distribution for the error variance. A better approach is to elicit enough assessments to give several estimates of the hyperparameters and to then reconcile these estimates in some way (Kadane and Wolfson, 1998). This same criticism applies to methods used to quantify experts' opinion about the variance of a multivariate normal distributions (Al-Awadhi and Garthwaite, 1998; 2001).

The purpose of this paper is to extend the method of Garthwaite and Dickey (1988) so as to obtain several estimates of the hyperparameter that is most difficult to assess (a degrees of freedom parameter). Reconciliation of these estimates (perhaps with further input from the expert) yields an overall estimate of it. Theory underlying the proposed method is given in Section 2 and implementation of the method is described in Section 3. The implementation uses interactive graphics and it forms part of an interactive graphics program for quantifying opinion about a generalised linear or piecewise-linear model (Garthwaite et al., 2011). However, the method proposed here could be used as an add-on to any method of quantifying opinion about the regression coefficients of a normal linear model.

2. The Assessment Method

We suppose a dependent variable Y is related to covariates $X_1, \ldots, X_m$ through the normal linear model.

$$Y = \alpha + \beta_1 X_1 + \cdots + \beta_m X_m + \varepsilon, \tag{1}$$

where ε is assumed to be a normal random error with zero mean and an unknown variance σ_ε^2, i.e.

$$\varepsilon \sim \mathrm{N}(0, \sigma_\varepsilon^2). \tag{2}$$

A conjugate prior distribution for σ_ε^2 is the inverted Chi-squared distribution [see, for example, Pratt, Raiffa and Schlaifer (1995), or Garthwaite and Dickey (1988)]. We assume,

$$\sigma_\varepsilon^2 \sim \text{Inverted Gamma}(\nu/2, \nu w/2), \tag{3}$$

so its probability density function is

$$f(\sigma_\varepsilon^2; \nu, w) = \frac{1}{\Gamma(\nu/2)} \left(\frac{\nu w}{2}\right)^{\nu/2} \left(\frac{1}{\sigma_\varepsilon^2}\right)^{(\nu/2)+1} \exp\left(-\frac{\nu w}{2\sigma_\varepsilon^2}\right),$$
$$\sigma_\varepsilon^2 > 0,\ \nu > 0,\ w > 0. \tag{4}$$

The parameters of this distribution (ν and w) are referred to as *hyperparameters* and are the quantities that must be assessed.

In the method of Garthwaite and Dickey (1988), the expert is asked to suppose that two responses, Y_1 and Y_2 say, are observed at the same values of the predictor variables, so that the difference between them is due only to random variation. Let $Z_1 = Y_1 - Y_2$. She is asked to give her median for their absolute difference, $|Z_1|$, and q_0 denotes her assessment. She is then asked to suppose that the actual difference between that pair of responses was a value the computer specifies, say z_1. Conditional on this hypothetical datum, she is then asked to suppose that a further pair of responses are observed at identical values of the predictor variables. (The actual values of the predictor variables are not specified as these do not matter, as long as they are the same for the two responses.) She gives her median for the absolute difference between this pair of responses and we denote her assessments by q_1. The hypothetical sample datum should cause her to revise her opinion; the degree of revision measures her confidence in her original assessment, yielding information about the degrees of freedom parameter, ν. Garthwaite and Dickey use the assessments q_0 and q_1 to estimate ν and w.

As noted in the introduction, using two assessments to determine two hyperparameters is undesirable. Here we extend this method so that a number of different hypothetical data sets are presented to the expert, the expert repeating the above task after each one. This yields several estimates of ν that can be averaged.

Each hypothetical data set consists of a number of pairs of observations, where the two observations in a pair are taken at the same values of the predictor variables. Focusing on any one hypothetical set, suppose it consists of k pairs of observations and let Z_i denote the difference between the ith pair. The expert is asked to suppose that $Z_1 = z_1, \ldots, Z_k = z_k$ and gives her conditional median for the absolute difference between a further pair of observations taken at one set of values of the predictor variables. This median is denoted q. We next derive formulae for estimating ν and w from q_0 and q.

Clearly, from (1) and (2), given σ_ε^2, the random variables $Z_1, \cdots, Z_k$ are independent and identically distributed normal variates, i.e. for $i = 1, 2, \cdots, k$,

$$Z_i|\sigma_\varepsilon^2 \sim \mathrm{N}(0, 2\sigma_\varepsilon^2), \tag{5}$$

with the joint distribution

$$f(z_1, \cdots, z_k|\sigma_\varepsilon^2) = \frac{1}{(4\sigma_\varepsilon^2\pi)^{k/2}} \exp\left(\sum_{i=1}^{k} z_i^2/4\sigma_\varepsilon^2\right), \qquad -\infty < z_i < \infty, \quad \sigma_\varepsilon^2 > 0. \tag{6}$$

From (4) and (6), the joint distribution of $Z_1, \cdots, Z_k$ and σ_ε^2 is given by

$$f(z_1, \cdots, z_k, \sigma_\varepsilon^2; \nu, w) = \frac{(\nu w/2)^{\nu/2}}{\Gamma(\nu/2)(4\pi)^{k/2}} \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{\nu+k}{2}+1} \exp\left\{-\frac{1}{4\sigma_\varepsilon^2}\left(\sum_{i=1}^{k} z_i^2 + 2\nu w\right)\right\}, \qquad -\infty < z_i < \infty, \quad \sigma_\varepsilon^2, \nu, w > 0. \tag{7}$$

Integrating σ_ε^2 out from the RHS of (7), we get

$$f(z_1, \cdots, z_k; \nu, w) = \frac{\Gamma((\nu+k)/2)}{\Gamma(\nu/2)[\nu\pi(2w)]^{k/2}} \left[1 + \frac{\sum_{i=1}^k z_i^2}{\nu(2w)}\right]^{-\dfrac{\nu+k}{2}},$$

$$-\infty < z_i < \infty, \quad \nu, w > 0. \tag{8}$$

which is the k-variate version of the general three-parameter Student-t distribution with $\nu + k$ degrees of freedom, zero mean vector and a diagonal scale matrix $2wI_k$, where I_k is the identity matrix of order k, i.e.

$$Z_1, \cdots, Z_k \sim \text{MV-}t_{\nu+k}(\underline{0}_k, 2wI_k). \tag{9}$$

Now, the conditional distribution of σ_ε^2 given $Z_1 = z_1, \cdots, Z_k = z_k$, can be obtained by dividing the RHS of (7) by that of (8) to get

$$f(\sigma_\varepsilon^2 | Z_1 = z_1, \cdots, Z_k = z_k; \nu, w) = \frac{1}{\Gamma((\nu+k)/2)} \left[\frac{1}{4}\left(2\nu w + \sum_{i=1}^k z_i^2\right)\right]^{(\nu+k)/2} \times$$

$$\left(\frac{1}{\sigma_\varepsilon^2}\right)^{\dfrac{\nu+k}{2}+1} \exp\left\{-\frac{1}{4\sigma_\varepsilon^2}\left[2\nu w + \sum_{i=1}^k z_i^2\right]\right\}, \qquad \sigma_\varepsilon^2, \nu, w > 0. \tag{10}$$

Since the inverted gamma distribution is a conjugate prior for σ_ε^2, comparing (10) with (4), we can write

$$(\sigma_\varepsilon^2 | Z_1 = z_1, \cdots, Z_k = z_k) \sim \text{Inverted Gamma}\left(\frac{\nu+k}{2}, \frac{(\nu+k)w_k}{2}\right), \tag{11}$$

where

$$w_k = \frac{1}{\nu+k}\left[\nu w + \frac{\sum_{i=1}^k z_i^2}{2}\right]. \tag{12}$$

If Z denotes the difference between two further observations taken at the same design point then $Z|\sigma_\varepsilon^2 \sim \text{N}(0, 2\sigma_\varepsilon^2)$, so the conditional distribution of $(Z|Z_1 = z_1, \cdots, Z_k = z_k)$ is given by

$$f(Z|Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)}) =$$

$$\int_{\sigma_\varepsilon^2=0}^{\infty} f(Z|\sigma_\varepsilon^2) \times f(\sigma_\varepsilon^2 | Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)}) d\sigma_\varepsilon^2. \tag{13}$$

The integrand in (13) is similar to the RHS of (7) with k set equal to 1, ν replaced by $\nu + k$, and w replaced by w_k.
As in (8) and (9), integrating σ_ε^2 out from (13) gives

$$(Z|Z_1 = z_1, \cdots, Z_k = z_k) \sim t_{\nu+k}(0, 2w_k). \tag{14}$$

Similarly, from (4) and (5), the marginal unconditional distribution of Z_1 is

$$Z_1 \sim t_\nu(0, 2w). \tag{15}$$

As q_0 is the assessed median of $|Z_1|$, it is also the upper quartile assessment of Z_1, since the distribution of Z_1 is symmetric about 0. Similarly, as q is the median of $|(Z|Z_1 = z_1, \cdots, Z_k = z_k)|$, it is also the upper quartile assessment of $(Z|Z_1 = z_1, \cdots, Z_k = z_k)$. If we denote the upper quartile of a standard Student-t distribution with n degrees of freedom by Q_n, then we have

$$q_0 = (2w)^{1/2} Q_\nu, \tag{16}$$

and

$$q = (2w_k)^{1/2} Q_{\nu+k}. \tag{17}$$

The aim now is to solve this pair of equations to obtain ν and w. By division,

$$\frac{q_0}{q} = \frac{Q_\nu}{Q_{\nu+k}} \left[\frac{w}{w_k}\right]^{1/2}. \tag{18}$$

Using (12) and (16) we can eliminate w from (18), giving

$$\frac{q_0}{q} = \frac{Q_\nu}{Q_{\nu+k}} \left[\frac{\nu + k}{\nu + Q_\nu^2 \sum_{i=1}^k (z_i/q_0)^2}\right]^{\frac{1}{2}}. \tag{19}$$

The implementation of the method uses a simple search procedure to find the value of ν that solves equation (19). For this approach to work, the function in (19) must be strictly monotonic in ν; we would like a condition on the z_i that gives this monotonicity.

Also, a solution to (19) must exist; this places restrictions on the value of q_0/q if these assessments are to be statistically coherent. The remainder of these section concerns these issues.

Let $C = \sum_{i=1}^{k}(z_i/q_0)^2$. Three-dimensional plots of $d(q_0/q)/d\nu)$ against ν and C (cf. Figure 1) showed that the RHS of (19) is strictly monotonic in of ν if

$$\frac{\sum_{i=1}^{k}(z_i/q_0)^2}{k} < 1.6. \tag{20}$$

As the z_i are chosen by the computer after the expert has assessed q_0, their values will satisfy (20).

It remains to specify limits on the assessment q. To this end, we assume there is a minimum reasonable value for the elicited degrees of freedom, say $min(\nu)$, and a maximum reasonable value, say $max(\nu)$. Since q_0 has already been assessed, using the two extreme values $min(\nu)$ and $max(\nu)$ of ν in the RHS of (19) gives the two boundaries of q, as follows

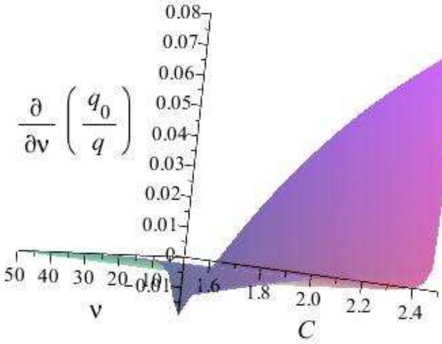
$$a = q_0 \frac{Q_{min(\nu)+k}}{Q_{min(\nu)}} \left[\frac{min(\nu) + Q^2_{min(\nu)} \sum_{i=1}^{k}(z_i/q_0)^2}{min(\nu)+k} \right]^{\frac{1}{2}}, \tag{21}$$

$$b = q_0 \frac{Q_{max(\nu)+k}}{Q_{max(\nu)}} \left[\frac{max(\nu) + Q^2_{max(\nu)} \sum_{i=1}^{k}(z_i/q_0)^2}{max(\nu)+k} \right]^{\frac{1}{2}}. \tag{22}$$

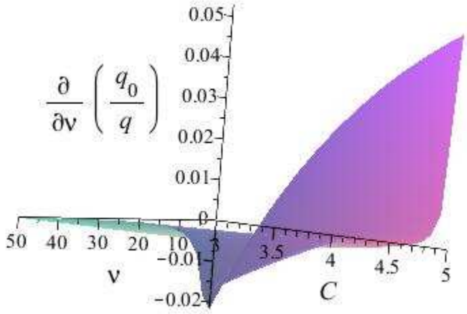
When the expert is making her assessments, we require the assessment q to be within the interval (a, b). Then the monotonicity of q_0/q as a function of ν ensures that there is a unique value of ν in the interval $(min(\nu), max(\nu))$ that satisfies (19). Given ν, equation (16) yields w.

2. Implementation

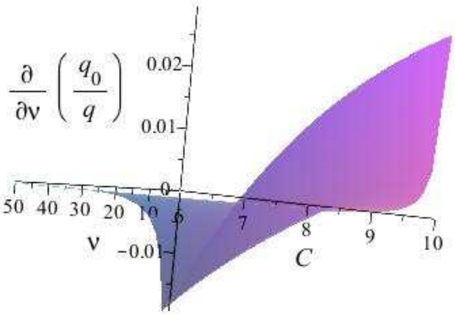
The elicitation method proposed in the previous section has been implemented as an interactive graphical procedure. The procedure could be used as a separate stand-alone program but it also forms part of interactive graphics software for quantifying opinion about a generalized linear or piecewise-linear model (Garthwaite et al., 2011). The



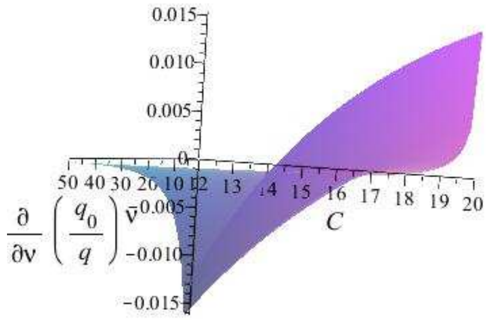
For $k = 1$, derivative negative for $C < 1.626$.



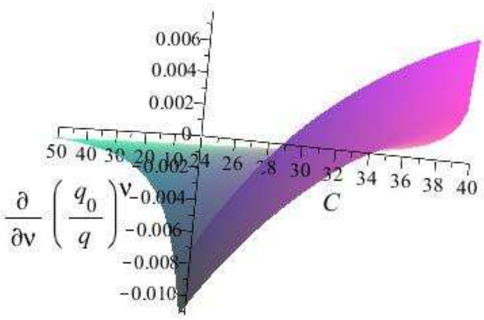
For $k = 2$, derivative negative for $C < 3.367$.



For $k = 3$, derivative negative for $C < 6.950$.



For $k = 4$, derivative negative for $C < 14.223$.



For $k = 5$, derivative negative for $C < 28.846$.

Figure 1. Three dimension plots of $\partial(q_0/q)/\partial\nu$ against ν and C for various sample sizes (k).

software gives an expert the option of assessing a prior distribution for the random error variance when an ordinary linear regression model is the GLM of interest. The software may be downloaded from http://statistics.open.ac.uk/elicitation.

The implementation first asks the expert to say whether the normal linear model relates to an experimental setting, observations on people, or observations on items. The aim is to frame questions in a way that is meaningful to the expert. In the context of an experiment, a set of values chosen for the predictor variables are referred to as a *design point*, so each pair of observations are *two responses at the same design point.* With observations on people (items), the observations are on *two people (items) whose covariate values give them identical characteristics (features).*

In a dialogue box, the expert is then asked to assume that two independent experiments have been conducted at the same design point. She assesses her median value of the absolute difference between the observed responses in these two virtual experiments. This assessed median is q_0.

Her remaining assessments are conditional assessments after being shown hypothetical data sets. The choice of the conditioning values in these data sets is an important issue. Garthwaite and Dickey (1988) note that hypothetical data should (a) be moderately different from the expert's initial beliefs, so that the data change the expert's beliefs by a measurable amount, but (b) should not be so different that the expert dismisses the data as being false and misleading. The hypothetical data set used by Garthwaite and Dickey (1988) consists of just one hypothetical difference, which they set at $q_0/2$ to try and meet these desiderata. Following Garthwaite and Dickey (1988) we take this as our first data set. This hypothetical datum is presented to the expert visually, using a diagram similar to that in Figure 2. In the figure, the value of the hypothetical datum is 1.5 and this is marked by the upward pointing arrow, the expert's original estimate of the median difference (q_0) is marked by the tall vertical line (at 3.0). The short lines below the horizontal axis mark the interval within which the expert's conditional median assessment of the difference must lie for statistical coherence. This interval was determined from equations (21) and (22) with $min(\nu)$ set equal to 1 and $max(\nu)$ set equal to 50. The expert makes her conditional assessment of the median difference by clicking the mouse-pointer on the horizontal axis within this interval. Let q_1 denote this assessment. Using equations (16) and (19), q_0 and q_1 yield the first estimates of ν and w, which we denote by ν_1 and w_1.

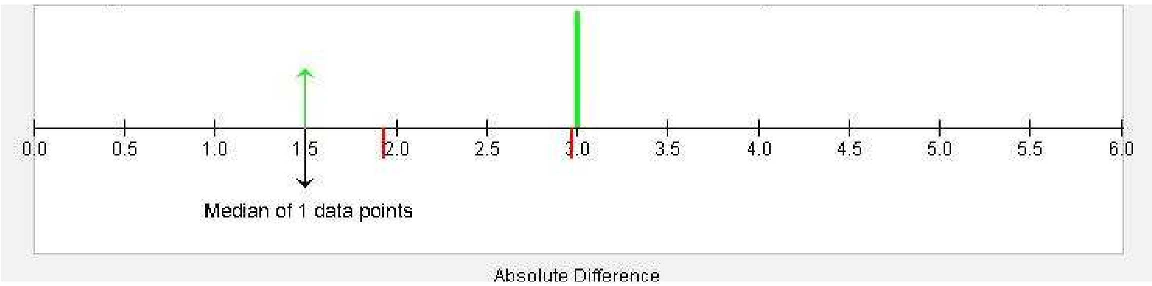


Figure 2. Elicitation of median assessment conditional on a single datum.

Our other hypothetical data sets contain more items. This gives these data sets greater weight and means that they should have greater effect on the expert's opinions, without the expert finding them unbelievable. In our past experience of using the method of Garthwaite and Dickey (1988), occasionally an expert has said that a single datum is too inconsequential to have any impact on his beliefs, which is also a problem that larger data sets should resolve. Kadane et al. (1980) present an expert with a sequence of hypothetical data sets and suggest that an expert should not be asked to forget any hypothetical data after it has been presented. They argue that "...asking the experimenter to forget would impose too great a psychological burden." (Kadane, et al., 1980, p. 849). In this spirit, we make our hypothetical data sets steadily bigger, so that each contains all the hypothetical data in its predecessors.

In our implementation we generate five hypothetical data sets. The jth set contains 2^{j-1} data ($j = 1, \ldots, 5$). A more flexible implementation would allow the expert to choose the number of data sets and the number of data in each, but it seems unlikely that a subject-matter expert would have knowledge relevant to making these choices. The first 2^{j-2} of these data $(z_1, \ldots, z_{2^{j-2}})$ come from the previous set of hypothetical data. The remaining items $(z_{2^{j-2}+1}, \ldots, z_{2^{j-1}})$ are obtained by generating random values from a normal distribution with zero mean and a variance of $(q_0/1.35)^2$. The z_i are set equal to the absolute values of the generated values. As 1.35 is the interquartile range of a standard normal distribution, the median of the z_i should about $q_0/2$. However, but random variation affects the z_i. the hypothetical data must be moderately different from the expert's initial beliefs, so as to give a measurable change in her opinions. In our implementation, $z_{2^{j-i}+1}, \ldots, z_{2^j}$ are resampled from the normal distribution if $\sum_{i=1}^{k} z_i^2/k > (3q_0/4)^2$. This also ensures that the constraint for monotonicity given by (20) will be satisfied.

After a suitable hypothetical data set has been generated, it is presented to the expert using the graphical interface. Figure 3 is an example in which a set of 16 hypothetical data is displayed (the arrows pointing up). Their median is the arrow pointing down and is a summary of the data that the expert may find helpful. The expert's original median assessment, q_0, is the rightmost tall vertical line. The other vertical line is the expert's median assessment conditional on the previous hypothetical data set. Taking account of all the hypothetical data, the expert gives her conditional assessment of the median absolute difference by clicking the mouse-pointer in the interval bounded by the short lines below the horizontal axis. One strategy for making this assessment is to consider the original median assessment and the median of the hypothetical data, and then choose a point between them that reflects their relative importance. For example, if the original median assessment is thought to be a better estimate than the median of the data, then the new median assessment should be nearer to the original median assessment than the median of the data. Let q_j denote the median assessment conditional on the jth set of hypothetical data.

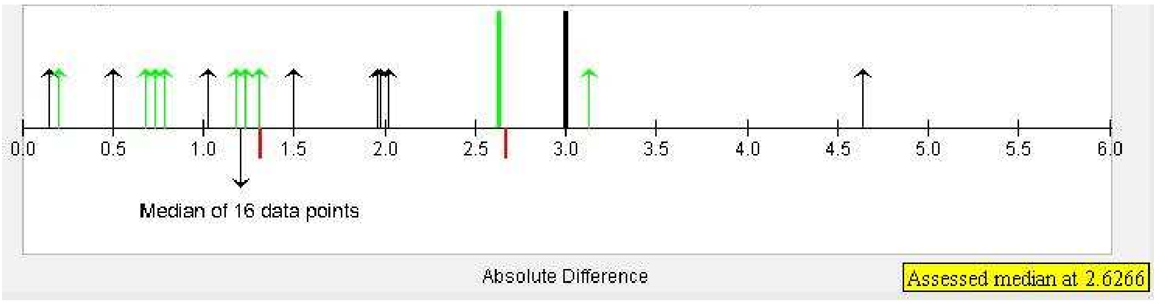


Figure 3. Elicitation of median assessment conditional on a set of data.

In conjunction with q_0, each q_j yields estimates of ν and w. Let ν_j and w_j denote the estimates for the jth data set ($j = 1, \ldots, 5$). The estimates are displayed in a table. In principle, the estimates of ν should all be very similar, as should the estimates of w. The expert is invited to revise her median assessment for any hypothetical data set (or sets) she wishes, and should take this option if some estimates appear out of line. When several estimates of a degrees of freedom parameter are to be quantified, empirical work suggests that it is better to take their geometric mean than their arithmetic mean (Al-Awadhi, 1997). Hence, when the estimates are acceptable to the expert, the geometric mean of $\nu_1, \ldots, \nu_5$, is taken as the value of ν in the prior distribution. Using the assessed value q_0, the corresponding value of w is determined from equation (16) and taken as the prior value of w.

References

Al-Awadhi, S. A. (1997). *Elicitation of Prior Distributions for a Multivariate Normal Distribution.* PhD. Thesis, University of Aberdeen.

Al-Awadhi, S. A. and Garthwaite, P. H. (1998). An elicitation method for multivariate normal distributions. *Communications in Statistics – Theory and Methods,* **27**, 1123–1142.

Al-Awadhi, S. A. and Garthwaite, P. H. (2001). Prior distribution assessment for a multivariate normal distribution: An experimental study. *Journal of Applied Statistics,* **28**, 5–23.

Alpert, M. and Raiffa, H. (1969). A Progress Report on the Training of Probability Assessors. In *Judgment Under Uncertainty: Heuristics and Biases,* D. Kahneman, P. Slovic, and A. Tversky (eds.), 294–305. Cambridge: Cambridge University Press.

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association,* **91**, 1450–1460.

Chen, M.-H., Ibrahim, J. G., Shao, Q.-M., and Weiss, R. E. (2003), "Prior Elicitation for Model Selection and Estimation in Generalized Linear Mixed Models," *Journal of Statistic Planning and Inference,* 111, 57–76.

Garthwaite, P. H., Al-Awadhi, S. A., Elfadaly, F., and Jenkinson, D. J. (2011). Prior distribution elicitation for generalized linear and piecewise-linear models. Submitted for publication.

Garthwaite, P. H. and Dickey, J. M. (1988). Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society,* **B50**, 462–474.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association,* **70**, 271–294.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S., and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association,* **75**, 845–854.

Kadane, J. B. and Wolfson, L. J. (1998). Experiences in elicitation. *The Statistician,* **47**, 3–19.

Pratt, J. W. and Raiffa, H. and Schalifer, R. (1995). *Introduction to Statistical Decision Theory.* London: MIT Press.