Elicitation of Prior Distributions For Variable-Selection Problems in Regression

Paul H. Garthwaite and James M. Dickey
University of Minnesota
School of Statistics
Technical Report No. 553
September, 1990

Research supported in part by UK Science and Engineering Research Council, Imperial Chemical Industries, and U.S. National Science Foundation (Grants MCS-830/335, DMS-8614793, DMS-8911548).

ELICITATION OF PRIOR DISTRIBUTIONS FOR VARIABLE-SELECTION PROBLEMS IN REGRESSION

BY PAUL H GARTHWAITE AND JAMES M DICKEY

University of Aberdeen and University of Minnesota

SUMMARY

This paper addresses the problem of quantifying expert opinion about a normal linear regression model when there is uncertainty as to which independent variables should be included in the model. Opinion is modelled as a mixture of natural conjugate prior distributions with each distribution in the mixture corresponding to a different subset of the independent variables. It is shown that for certain values of the independent variables, the predictive distribution of the dependent variable simplifies from a mixture of t-distributions to a single t-distribution. Using this result, a method of eliciting the conjugate distributions of the mixture is developed. The method is illustrated in an example.

AMS 1980 *subject classifications.* Primary 62F15; secondary 62C10, 62J05.

Key words and phrases. Probability assessment methods, probability elicitation, prior distribution, variable selection, linear regression.

14 September 1990.

1. Introduction. This paper is concerned with the task of quantifying an expert's opinion about a regression model when the expert is uncertain about which set of independent variables should be used in the model. It is supposed that a response, Y, is related to independent variables $X_1, \ldots, X_r$ through the usual normal sampling model

$$Y = \beta_1 X_1 + \ldots + \beta_r X_r + \varepsilon,$$

and the expert believes that one or more of the coefficients β_j are likely to be zero or trivially small. There are many situations of this form where it would be useful to have expert opinion expressed in a prior distribution. For example, motivation for the present work arose from the potential benefit of being able to use expert opinion in the design of experiments. At the design stage, the source of information is the experimenter's background knowledge, including information gained from previous experimental data. Also, at that stage, a variable-selection problem commonly arises because all the variables judged as having a nontrivial chance of a marked affect on the response should be included in the design. The failure to identify and control important variables could be a serious error. Questions of how to utilize prior distributions when designing experiments have been treated, for example, by Atkinson and Fedorov (1975a and 1975b).

Methods of quantifying subjective opinion about a linear regression model have been developed for the case where the variable-selection problem does not arise (e.g. Kadane *et al.,* 1980; Garthwaite and Dickey, 1988, 1990). Such methods assume that expert opinion can be well represented by a member of the standard family of conjugate prior distributions (Raiffa and Schlaifer, 1961), but this assumption may be inappropriate if the expert has prior suspicion that there may be X-variables included in the model that are unimportant. To illustrate, suppose that the response, Y, is the yield in an industrial chemical process and that X_j corresponds to the quantity of a chemical, where the chemical might be of a type that acts as a catalyst or might be one that has no effect. It follows that the expert's marginal prior distribution for β_j, $f(\beta_j)$ say, would include a sharp peak of probability at the origin, corresponding to the probability that X_j has virtually no effect. The remainder of the probability would be mainly to the right of the origin, corresponding to X_j being a catalyst and beneficial to the response. The distribution might then be similar to that illustrated in Fig. 1, which cannot be represented by the natural conjugate prior (a t-distribution).

Fig. 1. about here

It is imagined that if the effective-variable problem could be resolved, then opinion could be represented by a natural conjugate distribution. But since the subset of effective variables is not known, opinion will be represented by a mixture of conjugate distributions, where each constituent distribution corresponds to a different subset of regressor variables. A relationship between the constituent distributions of the mixture will be assumed that will result in the problem being tractable. The chosen relationship is described in the next section and gives a structure which permits marginal distributions of the type illustrated in Fig. 1, provided the sharp peak of probability can be well approximated by a point mass at the origin.

We give a method in this paper for eliciting the conjugate distribution constituents of the prior distribution, but we do not give a special method of eliciting the mixing weights, beyond asking directly for the subjective probabilities of possible sets of effective variables. The method given here is a generalisation of the conjugate-prior method of Garthwaite and Dickey (1988). Indeed, the method given in that paper [to be referred to here as G & D] was obtained as a special case during development of the method reported here. Both methods exploit an elicitation task involving the choice of points of Constrained Minimum Variance, or CMV points. In Section 3 this task is described and results developed concerning CMV points pertinent to the variable-selection problem. In Section 4 the elicitation method is described, and in Section 5 the way the elicited information is used to determine the conjugate distributions is given. An example illustrating the use of the method is provided in Section 6. The example also shows that assessing the mixing weights of the prior distribution can be straightforward.

The elicitation method has been implemented as an interactive computer programme. To quantify his or her opinion, the expert types in answers to questions displayed by the computer, questions formulated on the basis of the expert's answers to preceding questions. The individual assessment tasks he must perform are essentially similar to the tasks imposed in the elicitation method of G & D, despite the added complexity of having opinion modelled by a mixture of conjugate distributions rather than a single such distribution. A user guide for the computer programme, together with a programme listing and details of the implementation, are given in Garthwaite (1990). Further examples where the elicitation method has been used to quantify the opinions of industrial chemists may be found in Garthwaite (1983).

2. Model and notation. The sampling model states that the response Y is related to independent variables $X_1, \ldots, X_r$ by the equation

$$Y = \beta_1 X_1 + \ldots + \beta_r X_r + \varepsilon,$$

where the experimental error is ε and is normally distributed with mean 0 and (unknown) variance σ^2. We suppose $\beta_1 X_1$ is a constant term with X_1 identically equal to 1. It is also supposed that each independent variable can take on any value between its lower and upper bounds and that none of the variables are deterministically related to one another. Otherwise the CMV points would be excessively constrained, and as a consequence, would encode insufficient information (*c.f.* G & D). While some variables might not affect the response, there will usually be others which, in the expert's opinion, are certain to affect it. For convenience the variables are ordered so that the first m variables, $X_1(\equiv 1), X_2, X_3, \ldots, X_m$, $m \leq r$, are considered certain to affect the response.

Let $f(\underline{\beta}, \sigma)$ denote the expert's joint prior distribution for $\underline{\beta}$ and σ, where $\underline{\beta} = (\beta_1, \ldots, \beta_r)'$. The expert's opinion gives positive probability that some β-coefficients are zero. For each $i = 1, 2, \ldots, h$, let H_i be an hypothesis which specifies that certain β-coefficients are zero and that the other coefficients are non-zero. Also, let H_0 be the special hypothesis stating that all the β-coefficients are non-zero with probability one. It is assumed that exactly one of the $h+1$ hypotheses $H_0, H_1, \ldots, H_h$ is true and that each of these has positive probability of being true, with the possible exception of H_0. The prior distribution can then be expressed as a mixture of $h+1$ conditional distributions:

$$f(\underline{\beta}, \sigma) = \sum_{i=0}^{h} f(\underline{\beta}_{(i)}, \sigma \,|H_i)\, P(H_i) \tag{2.1}$$

where $\underline{\beta}_{(i)}$ denotes the non-zero β-coefficients when H_i is true. $P(H_i)$ is the expert's prior probability that H_i is the true hypothesis. Representing a prior distribution as a mixture of conditional distributions in this way has been advocated by Hill (1974), Dickey (1974, 1980), and others.

A relationship between the conditional distributions in (2.1) is required to make the elicitation problem tractable. One way of relating these distributions, a way that we will *not* use without modification, is first to take the distribution conditional on H_0 (H_0 gives zero probability that any β-coefficient is zero) and then to condition further on particular β-coefficients being zero. With each H_i $(i = 1, 2, \ldots, h)$, associate a set of integers, ρ_i say, for which $j \in \rho_i$ means that H_i requires β_j equal zero, and with probability one under H_i, the other β_j are non-zero. One might then assume the continuity condition,

$$f(\underline{\beta}_{(i)}, \sigma \,|H_i) = f(\underline{\beta}_{(i)}, \sigma \,|H_0, \beta_j = 0 \text{ for } j \in \rho_i). \tag{2.2}$$

Such prior continuity conditions are discussed generally by Dickey and Lientz (1970) and Gunel and Dickey (1974). They play an important role in Savage's density ratio for Bayes factors. Relationships of the form in (2.2) would arise, for example, if an expert were perfectly coherent in his opinions and all his knowledge of $\underline{\beta}$ and σ came from experiments with the regression model of current interest. That is, if each prior distribution under an hypothesis were non-informative, and sample data were then obtained, then the posterior distributions under the different hypotheses would satisfy equation (2.2).

A disadvantage of the structure given in (2.2) is that the marginal prior distribution of σ will vary from hypothesis to hypothesis. This would be inappropriate if an expert's opinions about the experimental error were mainly based, not on experimental work with the present problem, but on experience gained in other problems, perhaps using the same equipment or experimental techniques as will be required in the present problem. We believe that these latter circumstances occur commonly in practice.

In the case where σ is known, this disadvantage does not arise and the relationship derived by the further conditioning in (2.2) seems a suitable way to model expert opinion. Hence, we wish to choose a model that will have such a structure when σ is known, so we assume that

$$f(\underline{\beta}_{(i)}|H_i, \sigma) = f(\underline{\beta}\,|H_0, \sigma, \beta_j = 0 \text{ for } j \in \rho_i). \tag{2.3}$$

In the more general case where σ is unknown, the marginal distributions of σ conditional on the different hypotheses must also be specified, to define the joint distribution of $\underline{\beta}_{(i)}$ and σ conditional on H_i. In line with the observation in the preceding paragraph, we assume that this distribution is independent of which hypothesis is true. That is, for $i = 0, 1, ..., h$

$$f(\sigma) = f(\sigma\,|H_i). \tag{2.4}$$

Equations (2.3) and (2.4) give the relationships between the distributions in (2.1), since $f(\underline{\beta}_{(i)}, \sigma\,|H_i) = f(\underline{\beta}_{(i)}|H_i, \sigma)\, f(\sigma\,|H_i)$. Each distribution must also be given more specific structure. We suppose that each is a member of the natural conjugate family, as follows. Under every hypothesis, let σ^2 be distributed as ωn times the reciprocal of a chi-squared random variable with n degrees of freedom,

$$\sigma^2 \sim \omega n/\chi^2_n. \tag{2.5}$$

Given σ and H_0, let $\underline{\beta}$ have a normal distribution with some mean $\mathbf{b}$ and variance matrix $\sigma^2\mathbf{U}/\omega$. The distribution of $\underline{\beta}$, conditional on σ and any other hypothesis, is then given by (2.3) and is also multivariate normal. The hyperparameters in this prior distribution, ω, n, $\mathbf{b}$ and $\mathbf{U}$, together with the weights $P(H_i)$, must be determined in any elicitation method.

Conditional on any of the hypotheses H_i, the marginal distribution of $\underline{\beta}$ is a multivariate-t distribution with n degrees of freedom. The location-scale multivariate-t family with n degrees of freedom has a generic random vector $\mathbf{z} = \mathbf{c} + \mathbf{B}\mathbf{t}_n$, where $\mathbf{c}$ and $\mathbf{B}$ are constant and $\mathbf{t}_n$ is the standard multivariate-t vector on n degrees of freedom (Press, 1972). Following Kadane *et al.* (1980) and G & D, we define $C(\mathbf{z}) = \mathbf{c}$ as the "centre" of $\mathbf{z}$ and $S(\mathbf{z}) = \mathbf{B}\mathbf{B}'$ as the "spread" of $\mathbf{z}$. These quantities are used because they exist for all positive values of n, while the variance, $Var(\mathbf{z}) = [n/(n-2)]S(\mathbf{z})$, does not exist if n is less than 2 and the mean, $E(\mathbf{z}) = \mathbf{c}$, does not exist if n is less than 1. For $\underline{\beta}$, we have that $C(\underline{\beta}|H_0) = \mathbf{b}$ and $S(\underline{\beta}|H_0) = \mathbf{U}$.

3. Points of constrained minimum variance. A set of particular values $\mathbf{x}$ for the independent variables will be referred to as a design point, and $\overline{y}$ is used to denote the (unknown) average response that would be obtained if a large number of observations were obtained at a single design point. The main assessment tasks that the expert will perform in order to quantify his opinion are: (a) to select design points satisfying certain constraints, where subject to these constraints, his subjective accuracy in predicting $\overline{y}$ is maximised; and (b) to specify the median and quartiles of his predictive distribution for $\overline{y}$ at such points.

It has been assumed that the expert's prior distribution corresponds to a mixture of natural conjugate distributions, so the prior predictive distribution of $\overline{y}$ at (most) design points is a mixture of two or more t-distributions. Gaining useful information about a mixture distribution is a difficult task, since such quantities as its interquartile range bear no simple relationship to its parameters. To emphasize this point, three t-distributions and the mixture distribution they form are plotted in Fig. 2. It would clearly be difficult to obtain useful estimates of the parameters of the individual t-distributions through questioning the expert about the mixture. Instead, our approach is to find design points at which the prior predictive distribution simplifies from a mixture of distinct t-distributions to a single t-distribution. These points will be found to be points of constrained minimum variance, which we now define.

Fig. 2. about here

A *point of minimum variance* (MV) is a point where the interquartile range of the prior predictive distribution of $\overline{y}$ is minimized. It will be convenient to consider distributions conditional on H_i, and then to refer to an *MV point under* H_i. "Variance" is used in the terminology because, conditional on H_i, the predictive distribution of $\overline{y}$ at any design point is a t-distribution, so as $\mathbf{x}$ varies, $\mathrm{var}(\overline{y}\,|\,\mathbf{x}, H_i)$ (provided it exists) is proportional to the square of the interquartile range and both are minimised at the same design points.

Let $\mathbf{x}$ be partitioned so that $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')'$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ are $k \times 1$ and $(r-k)\times 1$ vectors, respectively, and suppose the constraint is imposed that $\mathbf{x}_1$ take some specified value, say $\mathbf{x}_1 = \mathbf{a}$. A point where the interquartile range of the predictive distribution of $\overline{y}$ is minimised, subject to this constraint, is referred to as a point of $\mathbf{a}$-*constrained minimum variance*, or if it is clear what constraint is meant, as just a *point of constrained minimum variance* (CMV). The MV point is a CMV point with k equal to 1 and $a = 1$, since X_1 is identically equal to 1 while the other X-variables are not constrained.

The basic result about CMV points (given by G & D, Theorem 4.1) is the following. Suppose $\mathbf{x}$ and $\mathbf{U}$ are conformably partitioned as

$$(3.1) \qquad \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \qquad \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix}.$$

Then if $\mathbf{x}_1$ is constrained to equal $\mathbf{a}$, the CMV point under H_0 is the point

$$(3.2) \qquad (\mathbf{a}', -\mathbf{a}'\mathbf{U}_{12}\mathbf{U}_{22}^{-1})'.$$

Also, the spread of the distribution of $\overline{y}$ at this point is given by

$$(3.3) \qquad S(\overline{y}\,|\,\mathbf{x}_1 = \mathbf{a}, \mathbf{x}_2, H_0) = \mathbf{a}'\mathbf{U}_{11.2}\mathbf{a}$$

where $\mathbf{U}_{11.2} = \mathbf{U}_{11} - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{U}_{21}$.

A CMV point under H_0 is unique. For a CMV point under other H_i, those X variables corresponding to non-zero β-coefficients are unique.

It will be convenient to express the above results in terms of inverse-spread matrices. Suppose,

$$\begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix}.$$

Then $-\mathbf{U}_{12}\mathbf{U}_{22}^{-1} = \mathbf{G}_{11}^{-1}\mathbf{G}_{12}$ and $\mathbf{U}_{11.2} = \mathbf{G}_{11}^{-1}$, so the $\mathbf{a}$-CMV point under H_0 is

$$(3.4) \qquad (\mathbf{a}', \mathbf{a}'\mathbf{G}_{11}^{-1}\mathbf{G}_{12})'$$

and

$$(3.5) \qquad S(\overline{y}\,|\,\mathbf{x}_1 = \mathbf{a}, \mathbf{x}_2, H_0) = \mathbf{a}'\mathbf{G}_{11}^{-1}\mathbf{a}.$$

It has been assumed to be known that $X_2, X_3, \ldots, X_m$ $(m \leq r)$ nontrivially affect the response. The following theorem shows that if some (or all) of these variables are constrained to take specified values, then the CMV point under H_0 is also a CMV point under every other H_i $(i = 1, \ldots, h)$, and at this point the distribution of $\overline{y}$ is a single t-distribution and not a more complicated mixture of t-distributions. We go on to show in Theorem 2 that this point is also the CMV point when it is uncertain which hypothesis is true. Proofs of the theorems are given in Appendix A.

THEOREM 1. *Let* $\mathbf{a} = (1, a_2, \ldots, a_k)'$ *where the* a_j *are constants and* $k \leq m$. *Then*

(i) *The* a*-CMV point under* H_0 *is also an* a*-CMV point under* H_i *for* $i = 1, 2, \ldots, h$.

(ii) *At this point* $\mathbf{x}$, $f(\overline{y} \,|\, \mathbf{x}) = f(\overline{y} \,|\, \mathbf{x}, H_i)$ *for* $i = 0, 1, \ldots, h$ *and* $f(\overline{y} \,|\, \mathbf{x})$ *is a t-distribution.* □

THEOREM 2. *Let* $\mathbf{a} = (1, a_2, \ldots, a_k)'$ *where the* a_j *are constants and* $k \leq m$. *Then the* a*-CMV point, when it is uncertain which hypothesis is true, is the* a*-CMV point under* H_0. □

The predictive distribution of $\overline{y}$ is a *mixture* distribution, so its variance at different design points is not proportional to its interquartile range. Since CMV points are defined in terms of the interquartile range of the distribution of $\overline{y}$, Theorem 2 does not show that var$(\overline{y} \,|\, \mathbf{x})$ is smaller at the a-CMV point than at any other point whose first components equal **a**. However, this result does hold, as given in the following theorem.

THEOREM 3. *If* $n > 2$ *(so that var*$(\overline{y} \,|\, \mathbf{x})$ *exists), then under the conditions of Theorem 2, var*$(\overline{y} \,|\, \mathbf{x})$ *is smaller at the* a*-CMV point than at any other point whose first* k *components equal* **a**.

Proof. *Let θ be a random variable that takes the value i if H_i is the hypothesis that is true. Then, for fixed* $\mathbf{x}$,

$$\begin{aligned} \mathrm{Var}(\overline{y}\,|\,\mathbf{x}) &= E_\theta[\mathrm{var}(\overline{y}\,|\,\mathbf{x}, H_\theta)] + \mathrm{Var}_\theta[E(\overline{y}\,|\,\mathbf{x}, H_\theta)] \\ &= \sum_{i=0}^{h} P(H_i)\mathrm{var}(\overline{y}\,|\,\mathbf{x}, H_i) + \mathrm{Var}_\theta[E(\overline{y}\,|\,\mathbf{x}, H_\theta)]. \end{aligned} \tag{3.6}$$

At the CMV point, $E(\overline{y}\,|\,\mathbf{x}, H_i) = E(\overline{y}\,|\,\mathbf{x})$ for all i (Theorem 1), so $\mathrm{Var}_\theta[E(\overline{y}\,|\,\mathbf{x}, H_\theta)] = 0$ at this point. Also, for all i, $\mathrm{var}(\overline{y}\,|\,\mathbf{x}, H_i)$ is smaller at this point than at any other point whose first k components equal $\mathbf{a}$. (Theorem 1). Hence (3.6) is also smaller at this point than at other points satisfying the constraint. □

The purpose of this section was to identify points at which the prior predictive distribution of $\overline{y}$ is a single t-distribution. Theorems 1 and 2 show that particular CMV points have this property. Moreover, the decisiveness with which $\mathrm{var}(\overline{y}\,|\,\mathbf{x})$ is minimised at such points (*each term* on the right-hand side of (3.6) is individually minimised) suggests that assessing the positions of CMV points is a reasonable task to ask of an assessor.

4. Elicitation method. In the prior model, the marginal distribution of σ satisfies $f(\sigma) = f(\sigma\,|H_i)$ for all i (equation 2.4) and its form is given in equation (2.5). To determine ω and n, the hyperparameters of this marginal distribution, the procedure given by G & D can be used without change. It is outlined briefly in Appendix B but is not discussed further in this paper. Instead, attention is concentrated on the other hyperparameters to be determined, $\mathbf{b}$ and $\mathbf{U}$. These latter parameters are the centre and the spread of the conditional prior distribution $f(\underline{\beta}\,|H_0,\sigma)$, so information about this distribution must be elicited. It might seem natural to ask conditional questions of the form: *"Suppose H_0 were true, what would be your assessment of ..."*. However, conditional questions are harder to answer than unconditional questions and become harder as the number of conditions increase. Also, a conditional question is particularly hard to answer when the given condition seems unrealistic, and '*H_0 is true*' may be such a condition. If several of the independent variables are each unlikely to affect the response, then H_0, the hypothesis that all variables affect the response, may be very unlikely. For these reasons, we prefer questions in which no hypothesis is conditioned on, and when conditional questions are asked, relatively weak conditions will be specified.

The elicitation method described here and the method of G & D require similar tasks of the expert. The positions of CMV points are elicited, together with fractile assessment of the predictive distribution of $\overline{y}$ at these points. The only difference is that in the present method, the expert is asked to assume, for parts of the elicitation interview, that a specified X variable is certain to affect the response. This is done for various X-variables in turn. For these parts of the interview, the expert must then take this assumption into account when giving assessments. To illustrate, if X corresponds to a variable which may, in the expert's opinion, be a catalyst in a chemical reaction, then at times the expert must assume that it is a catalyst when giving his assessments.

4.1. *Design point assessments.* Theorem 2 indicates that the overall MV point and the MV point *under* H_0 are coincident. Hence the questions (a) *"What is the MV point?"* and (b) *"Suppose H_0 were true, what would be the MV point?"* should, in principle, give the same answer. Consequently, to obtain the answer to the conditional question (b), the unconditional question (a) can be asked. Similarly, if $\mathbf{a} = (1, a_2, \ldots, a_k)'$ with $k \leq m$, the expert can be asked to specify the a-CMV point and his answer can then be equated to the a-CMV point *under* H_0. (The expert believes the first m X-variables are certain to affect the response, and is only uncertain as to which of the last $n - m$ variables should feature in the model.) Hence, without asking conditional questions, design points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ can be elicited that have the following structure:

$$\begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_j' \\ \vdots \\ \mathbf{x}_m' \end{pmatrix} = \begin{pmatrix} 1 & x_{1,2} & x_{1,3} & \cdots & \cdots & \cdots & x_{1,m} & x_{1,m+1} & \cdots & x_{1,r} \\ 1 & a_2 & x_{2,3} & \cdots & \cdots & \cdots & x_{2,m} & x_{2,m+1} & \cdots & x_{2,r} \\ \vdots & \vdots & & & & & \vdots & \vdots & & \vdots \\ 1 & a_2 & \cdots & a_j & x_{j,j+1} & \cdots & x_{j,m} & x_{j,m+1} & \cdots & x_{j,r} \\ \vdots & \vdots & & & & & & \vdots & & \vdots \\ 1 & a_2 & \cdots & \cdots & \cdots & \cdots & a_m & x_{m,m+1} & \cdots & x_{m,r} \end{pmatrix} \tag{4.1}$$

where:

(i) $\mathbf{x}_1$ is the MV point under H_0;

(ii) for $j = 2, \ldots, m$, $\mathbf{x}_j$ is the $\mathbf{a}_j$-CMV point under H_0, where $\mathbf{a}_j = (1, a_2, \ldots, a_j)'$;

and

$$a_j \neq x_{j,j-1} \qquad \text{for } j = 2, 3, \ldots, m. \tag{4.2}$$

In the above matrix, the values a_k are specified by the computer, and the values of the elements $x_{i,j}$ are chosen by the expert. The rows of the matrix are determined sequentially, starting with $\mathbf{x}_1$, which is obtained by eliciting the expert's MV-point. For $\mathbf{x}_j$ ($j = 2, \ldots, m$), the computer selects a value for a_j that differs from $x_{j,j-1}$, the jth element of the *preceding* row, thus satisfying (4.2). The other elements of $\mathbf{a}_j$ have previously been selected and the expert assesses his $\mathbf{a}_j$-CMV point, giving the point $\mathbf{x}_j$ specified in (ii).

To obtain the positions of further CMV points under H_0 , conditional questions are asked. For each $j = m+1, \ldots, r$, it is uncertain whether the variable X_j will have an affect on the response. For each of these variables in turn, the expert is asked to assume that it does have an effect. Conditional on this assumption, he assesses the CMV point for the constraint that (a) the first m components of the point equal $1, a_2, \ldots, a_m$ (these are the values of the first m components of $\mathbf{x}_m$) and (b) the jth component of the point equals a_j. Theorem 2 implies that, conditional on X_j affecting the response, the selected CMV point is also the CMV point under H_0. In this way, CMV points $\mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \ldots, \mathbf{x}_r$ are elicited that have the following form:

$$\begin{pmatrix} \mathbf{x}'_{m+1} \\ \mathbf{x}'_{m+2} \\ \vdots \\ \mathbf{x}'_j \\ \vdots \\ \mathbf{x}'_r \end{pmatrix} = \begin{pmatrix} 1 & a_2 \ldots a_m & a_{m+1} & x_{m+1,m+2} & & \cdots & \cdots & \cdots & & x_{m+1,r} \\ 1 & a_2 \ldots a_m & x_{m+2,m+1} & a_{m+2} & x_{m+2,m+3} & & \cdots & \cdots & & x_{m+2,r} \\ \vdots & \vdots \quad \vdots & \vdots & & & & & & & \vdots \\ 1 & a_2 \ldots a_m & x_{j,m+1} & \cdots\cdots\cdots & & x_{j,j-1} & a_j & x_{j,j+1} & \cdots & x_{j,r} \\ \vdots & \vdots \quad \vdots & \vdots & & & & & & & \\ 1 & a_2 \ldots a_m & x_{r,m+1} & \cdots\cdots\cdots & & \cdots & \cdots & \cdots & x_{r,r-1} & a_r \end{pmatrix} \tag{4.3}$$

The values $a_{m+1}, a_{m+2}, \ldots, a_r$ are selected by the computer and satisfy

$$a_j \neq x_{m,j} \qquad \text{for } j = m+1, \ldots, r. \tag{4.4}$$

4.2. *Median and quartile assessments.* At each of the design points $\mathbf{x}_1, \ldots, \mathbf{x}_r$, the expert is asked to assess the median, upper and lower quartiles of the predictive distribution, $f(\overline{y}\,|\mathbf{x}_j)$. For the point $\mathbf{x}_j$, where $j = m+1, m+2, \ldots, r$, he is asked to assume that the independent variable X_j certainly affects the response when making these assessments. Under this condition, Theorem 1 implies that $f(\overline{y}\,|\mathbf{x}_j)$ is identical to $f(\overline{y}\,|\mathbf{x}_j, H_0)$, and so the fractile assessments can be equated to fractiles of this latter distribution. Let $\overline{y}_{j,.50}$, $\overline{y}_{j,.75}$ and $\overline{y}_{j,.25}$ denote the median and quartile assessments at the point $\mathbf{x}_j$. Then for $j = 1, \ldots, r$, the centre and spread of $f(\overline{y}\,|\mathbf{x}_j, H_0)$ are calculated as

$$C(\overline{y}\,|\,\mathbf{x}_j, H_0) = \overline{y}_{j,.50} \tag{4.5}$$

$$S(\overline{y}\,|\,\mathbf{x}_j, H_0) = [(\overline{y}_{j,.75} - \overline{y}_{j,.25})\,/\,(2q_n)]^2 \tag{4.6}$$

where q_n is the interquartile range of a t-distribution with unit spread and n degrees of freedom.

As in G & D, the expert is also questioned about the differences in average response between pairs of design points. Specifically, the median and quartiles of the distributions of $d_2, d_3, \ldots, d_r$ are elicited, where $d_j = (\overline{y}\,|\mathbf{x}_j) - (\overline{y}\,|\,\mathbf{x}_s)$, and s is the smaller of $j-1$ and m. The usefulness of these assessments stems from results in the following theorem.

THEOREM 4. *For* $j = 2, 3, \ldots, r$

$$C(d_j\,|H_0) = C(\overline{y}\,|\,\mathbf{x}_j, H_0) - C(\overline{y}\,|\,\mathbf{x}_s, H_0) \tag{4.7}$$

$$S(d_j\,|H_0) = S(\overline{y}\,|\,\mathbf{x}_j, H_0) - S(\overline{y}\,|\,\mathbf{x}_s, H_0) \tag{4.8}$$

and the distribution of $f(d_j)$ is a t-distribution that is identical to $f(d_j \,|H_0)$.

Proof. For $i = 0, 1, \ldots, h$, trivially $C(d_j \,|H_i) = C(\overline{y}\,|\mathbf{x}_j, H_i) - C(\overline{y}\,|\,\mathbf{x}_s, H_i)$. From Theorem 1, $C(\overline{y}\,|\mathbf{x}_j, H_i)$ is the same for all H_i, so the same is true of $C(d_j \,|H_i)$. G & D show that $S(d_j \,|H_i) = S(\overline{y}\,|\,\mathbf{x}_j, H_i) - S(\overline{y}\,|\,\mathbf{x}_s, H_i)$ and, again from Theorem 1, $S(\overline{y}\,|\mathbf{x}_j, H_i)$ is the same for all H_i. Hence $S(d_j \,|H_i)$ is the same for all H_i. Clearly, for all H_i, $f(d_j \,|H_i)$ is a t-distribution on n degrees of freedom and we have just established that its centre and spread do not change as i varies. It follows that the distributions $f(d_j \,|H_i)$ are identical for $i = 0, 1, \ldots, h$, and hence equal $f(d_j)$. □

The theorem implies that median and quartile assessments of $f(d_j)$ can be equated to the corresponding fractiles of $f(d_j \,|H_0)$. The median is $C(d_j \,|H_0)$ and, analogous to (4.6), $S(d_j \,|\, H_0)$ is set equal to $[(d_{j,.75} - d_{j,.25})/(2q_n)]^2$. In the elicitation method, the expert is questioned about both $\overline{y}\,|\mathbf{x}_j$ and d_j at each design point in turn, and medians and quartiles of their distributions are elicited that give centres and spreads which satisfy equations (4.7) and (4.8). The expert is helped in this task by the computer. The expert assesses fractiles for d_j [or $(\overline{y}\,|\mathbf{x}_j)$] and the computer calculates fractiles for $\overline{y}\,|\mathbf{x}_j$ [or d_j] that would be consistent with these assessments. The expert then either accepts the calculated fractile values as being an adequate representation of his opinions, or he revises them. In the latter case the cycle is repeated. Requiring equations (4.7) and (4.8) to hold makes the expert consider the coherence between his assessments at different design points.

Equation (4.8) implies that, for $j = 2, \ldots, r$, the interquartile range for $\overline{y}$ should be smaller at $\mathbf{x}_s$ than at $\mathbf{x}_j$. This conclusion can also be drawn from the fact that, when $\mathbf{a} = (1, a_2, \ldots, a_s)$, $\mathbf{x}_s$ was chosen as the a-CMV point and not $\mathbf{x}_j$, even though the latter point also satisfies the constraint. (From (4.2) and (4.4), $\mathbf{x}_j \neq \mathbf{x}_s$.) If the interquartile range for $\overline{y}$ is not less at $\mathbf{x}_s$ than at $\mathbf{x}_j$, the expert is required to revise some of his previous fractile assessments and/or the positions of CMV points.

5. Assessment of hyperparameters. The elicited centres and spreads of predictive distributions, together with the elicited positions of CMV points, must be used to determine the hyperparameters $\mathbf{U}$ and $\mathbf{b}$ of the conditional prior distribution $f(\underline{\beta}\,|H_0)$.

5.1. *Assessment of* $\mathbf{U}$. Let $\mathbf{z}_i$ be the r-dimensional vector whose ith component equals 1 and whose other components are zero. Define the triangular matrix $\mathbf{T}$ by

(5.1) $\mathbf{T}' = (\mathbf{x}_1, \mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_2, \ldots, \mathbf{x}_m - \mathbf{x}_{m+1}, \mathbf{z}_{m+1}, \mathbf{z}_{m+2}, \ldots, \mathbf{z}_r)$.

Results given in G & D (Lemma 5.2 and Theorem 5.1) indicate that

$$S(\mathbf{T}\underline{\beta}\,|H_0) = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix} \tag{5.2}$$

where $\mathbf{D}$ is a $m \times m$ diagonal matrix whose non-zero elements are $S(\overline{y}\,|\mathbf{x}_1, H_0)$, $S(d_2\,|H_0)$, $S(d_3\,|H_0)$, ..., $S(d_m\,|H_0)$ and

$$\mathbf{V} = S((\beta_{m+1}, \beta_{m+2}, \ldots, \beta_r)'\,|H_0). \tag{5.3}$$

From (4.1), the diagonal elements of $\mathbf{T}$ are 1, $a_2 - x_{1,2}$, $a_3 - x_{2,3}$, ..., $a_m - x_{m-1,m}$, 1, 1, ..., 1. These are all non-zero, from (4.2), so $\mathbf{T}$ is invertible.

The expert's quartile assessments provide estimates of the non-zero elements of **D** so, if **V** can be estimated, then $\mathbf{U} = S(\underline{\beta}\,|H_0)$ can be calculated from

(5.4) $$\mathbf{U} = \mathbf{T}^{-1}.[S(\mathbf{T}\underline{\beta}\,|H_0)].(\mathbf{T}')^{-1}.$$

Denote $\mathbf{V}^{-1}$

$$\mathbf{V}^{-1} = \begin{pmatrix} g_{m+1,m+1} & g_{m+1,m+2} & \cdots\cdots & g_{m+1,r} \\ g_{m+2,m+1} & g_{m+2,m+2} & \cdots\cdots & g_{m+2,r} \\ \vdots & \vdots & & \vdots \\ g_{r,m+1} & g_{r,m+2} & \cdots\cdots & g_{r,r} \end{pmatrix}.$$

In Appendix C we show that the diagonal elements of this matrix may be estimated from

(5.5) $$g_{j,j} = (a_j - x_{m,j})^2 / S(d_j\,|H_0)$$

and that the off-diagonal elements should satisfy

(5.6) $$g_{j,i} = g_{j,j}(x_{j,i} - x_{m,i}) \,/\, (a_j - x_{m,j}).$$

Since the matrix $\mathbf{V}^{-1}$ is symmetric, the values of $g_{j,i}$ and $g_{i,j}$ given by (5.6) should be equal. To reconcile any difference, for simplicity we take their average as the estimate of $g_{i,j}$:

(5.7) $$g_{j,i} = \frac{1}{2}\left(\frac{g_{j,j}(x_{j,i} - x_{m,i})}{a_j - x_{m,j}} + \frac{g_{i,i}(x_{i,j} - x_{m,j})}{a_i - x_{m,i}}\right).$$

In the implementation of the elicitation method, the matrix

$$\begin{pmatrix} g_{m+1,m+1} & \cdots & g_{m+1,j} \\ \vdots & & \vdots \\ g_{j,m+1} & \cdots & g_{j,j} \end{pmatrix} \tag{5.8}$$

is estimated after assessments at the design point $\mathbf{x}_j$ have been elicited ($j = m+1, ..., r$). It is checked that this matrix is positive-definite, since otherwise the expert's assessments would not be probabilistically coherent. If this check were not satisfied, the expert would be required to revise some of his assessments. [In the authors' experience, these checks have always been satisfied]. A re-assessment procedure is given in Garthwaite (1990). When $j = r$, the matrix in equation (5.8) is $\mathbf{V}^{-1}$, so both $\mathbf{V}^{-1}$, and hence $\mathbf{V}$, will be positive-definite. After determining $\mathbf{V}$, the hyperparameter $\mathbf{U}$ is obtained from equations (5.1), (5.2) and (5.4), and it will also be positive-definite.

5.2. *Assessment of* **b**. The expert has given assessments that equate to $C(\overline{y}\,|\,\mathbf{x}_1, H_0)$, $C(d_2\,|H_0)$, ..., $C(d_r\,|H_0)$, where $C(d_j\,|H_0) = C(\overline{y}\,|\,\mathbf{x}_j, H_0) - C(\overline{y}\,|\,\mathbf{x}_s, H_0)$ and s is the smaller of $j-1$ and m. Let $\mathbf{d}_{.50} = (C(\overline{y}\,|\,\mathbf{x}_1, H_0),$ $C(d_2\,|H_0), \ldots, C(d_r\,|H_0))'$ and define the matrix $\mathbf{A}$ by $\mathbf{A}' = (\mathbf{x}_1, \mathbf{x}_2-\mathbf{x}_1, \ldots,$ $\mathbf{x}_m-\mathbf{x}_{m-1}, \mathbf{x}_{m+1}-\mathbf{x}_m, \mathbf{x}_{m+2}-\mathbf{x}_m, \ldots, \mathbf{x}_r-\mathbf{x}_m)$. We have that $\mathbf{d}_{.50} = \mathbf{Ab}$ and, in Appendix C, we show that the positive-definiteness of $\mathbf{V}$ ensures that $\mathbf{A}$ is non-singular. Thus the hyperparameter $\mathbf{b}$ can be determined as

$$\mathbf{b} = \mathbf{A}^{-1}\mathbf{d}_{.50}. \tag{5.9}$$

6. An example. In this real example, the "expert" whose opinion was quantified was an industrial chemist. He was seeking a viable way to manufacture a particular chloride compound. To produce this compound, two gasses are mixed in a diluent and passed through a long tube containing a catalyst. To the extent that the desired reaction does not occur, a waste product is produced and the chemist wanted to minimise the proportion of this waste product in the output. He was sure it would be affected by the following four factors: the temperature within the tube (*Temp 1*), the time the gas is in contact with the catalyst (*Time*), and the quantity of each gas (*Gas 1* and *Gas 2*) per unit volume of diluent. The chemist was also interested in three further factors which he thought might (but might not) affect the percentage waste: the temperature of the input gasses (*Temp 2*), the pressure (*Pres*) and the back-mix temperature (*Temp 3*). The chemist thought that those factors which affected the percentage waste would have a linear effect for the range of values he wished to consider. Hence if all factors had non-zero affects, the linear regression model for this application would be

$$\text{Waste} = \beta_1 + \beta_2(\textit{Temp 1}) + \beta_3(\textit{Time}) + \beta_4(\textit{Gas 1}) + \beta_5(\textit{Gas 2}) + \beta_6(\textit{Temp 2}) + \beta_7(\textit{Pres}) + \beta_8(\textit{Temp 3})$$

Before having his opinion elicited, the chemist was forewarned of the elicitation questions he would be asked and some advice was given on how he might tackle the questions. He had used an earlier version of the method so this took little time. The interactive computer program that implements the method was then initiated. In response to prompts from the computer the chemist typed in answers expressing his opinions.

His first set of answers determined the names and ranges of the independent variables that he felt certain would affect the response. These were:

Temp 1 :	360–445 ($°C$)	*Time* :	4–20 (secs)
Gas 1 :	5–15 (%)	*Gas 2* :	5–11 (%)

His next answers described the other variables which he thought might have an effect:

Temp 2 :	300–420 ($°C$)	*Pres* :	0-1 (atm)
Temp 3 :	250–380 ($°C$)		

[The chemist specified pressure as the increase in pressure above one atmosphere, measured in atmospheres.] He was next questioned about experimental error (using the methods of G & D), and his assessments gave values of 63.3 and 7 for ω and n, respectively.

The chemist then assessed the position of constrained points of minimum variance and quartiles of corresponding $\overline{y}$ and d. The co-ordinates of the selected points are given in Table 1. The values with an asterisk were chosen by the computer and the remainder were chosen by the assessor. The matrix

Table 1. about here

$\mathbf{T}$ defined in equation (5.1) is thus equal to

$$\begin{pmatrix} 1 & 380 & 8 & 9 & 6 & 320 & 0 & 280 \\ 0 & 22.5 & 0 & 0 & 1 & 30 & 0 & 0 \\ 0 & 0 & 8 & 0 & 2 & -30 & 0 & 0 \\ 0 & 0 & 0 & 3.5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 10 & 0 & -40 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The quartile assessments of $\overline{y}$ and d at the design points are given in Table 2. Values with an asterisk were suggested by the computer and accepted by the chemist as representative of his opinions. Only for the point $\mathbf{x}_8$ did the expert change a value ($d_{.50}$) that the computer suggested.

Table 2. about here

The semi-interquartile range of a standard t-distribution with 7 degrees of freedom is 0.711. From equations (4.6) and (4.8), the quartile assessments give the following respective values for $S(\overline{y}\,|\,\mathbf{x}_1, H_0), S(d_2\,|H_0), \ldots, S(d_5\,|H_0)$: 40.06, 4.45, 17.80, 17.80 and 4.45. These are the non-zero elements of the diagonal matrix $\mathbf{D}$, defined in equation (5.2).

The matrix V^{-1} is obtained from assessments at design points $\mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7$, and $\mathbf{x}_8$. Applying equations (5.5) and (5.7) yields

$$\mathbf{V}^{-1} = \begin{pmatrix} 808.8 & 1.123 & 842.5 \\ 1.123 & 0.05617 & 2.247 \\ 842.5 & 2.247 & 2477.0 \end{pmatrix}.$$

Inverting this matrix gives the remaining elements of $S(\mathbf{T}\underline{\beta}|H_0)$ [*c.f.* equation (5.2)]. The hyperparameter $\mathbf{U} = S(\underline{\beta}|H_0)$ is then obtained from equation (5.4), and equals

Const.	*Temp 1*	*Time*	*Gas 1*	*Gas 2*	*Temp 2*	*Pres*	*Temp 3*
2112	−5.70	−6.91	−16.5	30.2	.763	3.09	−.459
−5.70	.0179	.0113	.0103	−.0985	−.00356	.00875	.00156
−6.91	.0113	.360	.0683	−.301	.00160	−.0944	.00156
−16.5	.0103	.0683	1.56	−.259	.00092	.0896	−.00270
30.2	−.0985	−.301	−.259	1.54	.0225	.186	−.0159
0.763	−.00356	.00160	.00092	.0225	.00192	−.0128	−.00064
3.09	.00875	−.0944	.0896	.186	−.0128	18.6	−.0125
−0.459	.00156	.00156	−.00270	−.0159	−.00064	−.0125	.000634

The hyperparameter **b** is obtained from the assessed medians and coordinates of the design points. Applying equation (5.9) gives

$$\mathbf{b}' = (111.7, -0.120, -0.88, -1.75, 1.59, -0.029, -5.9, -0.012).$$

After the interactive elicitation interview, the chemist was given an explanation of the implications of the derived hyperparameter values that defined his assessed distribution. He thought the regression coefficient estimates represented his opinions quite well but, as one might have expected, the derived value of the spread matrix **U** meant little to him.

To complete the specification of the prior distribution for the linear model, mixing weights must be determined [the $P(H_i)$ in equation (2.1)]. The independent variables that might have no effect on the response are *Temp 2*, *Pres* and *Temp 3*. In discussion, the chemist responded to straightforward questions by asserting a one-in-five chance that *Temp 2* would affect the response, and for each of *Pres* and *Temp 3*, he assessed the probability at 0.1. Also, he felt that if *Temp 2* did affect the response, there was a probability of 0.2 that *Temp 3* would, as well. Knowing whether *Pres* affected the response would not change his probabilities of *Temp 2* or *Temp 3* affecting the response. These assessments enable the $P(H_i)$ to be determined. $P(H_0) = 0.004$ and, for $\rho_i = \{6,7,8\}$, $\{7,8\}$, $\{6,7\}$, $\{7\}$, $\{6,8\}$, $\{8\}$ and $\{6\}$, the corresponding $P(H_i)$ equal 0.666, 0.144, 0.054, 0.036, 0.074, 0.016 and 0.006, respectively.

7. Concluding remarks. For a variable-selection problem, subjective opinion should be modelled by a mixture of distributions, and structure can be imposed on the relationship between these distributions to reduce the hyperparameters that must be elicited to a manageable number. The structure adopted here seems sensible and a natural one to choose. With many forms of mixture distributions, eliciting the parameters of the individual distributions could be a formidable task. However, the properties of CMV points make the elicitation task reasonably straightforward for the model chosen here to represent subjective opinion. The assessment tasks that the expert must perform are only marginally more complicated than those required to determine a single conjugate distribution, rather than a mixture. The only difference is that, here, the expert must assume, in turn, that each of the independent variables is certain to affect the response. The calculations to form a prior distribution from the expert's assessments are somewhat more complicated than in G & D, but this is inevitable if one is to avoid asking the expert to make assumptions that are very unlikely or even impossible, such as 'H_0 *is true*'. We have sought to make the assessor's task as simple as possible, regardless of added complexity in the calculations. The example in the preceding section here and experiments reported elsewhere (Garthwaite, 1983, pp 130-136) indicate that the elicitation method developed is a usable procedure for quantifying expert opinion.

APPENDIX A

PROOF OF THEOREM 1. Without loss of generality, suppose H_i specifies that the last n_i components of $\underline{\beta}$ are zero and the first $r - n_i$ components are non-zero with probability one. Conformably partition $\mathbf{G} = [S(\underline{\beta}\,|H_0)]^{-1}$, $\underline{\beta}$ and $\mathbf{b}$ as follows:

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{W}_{12} & \mathbf{W}_{13} \\ \mathbf{W}_{21} & \mathbf{W}_{22} & \mathbf{W}_{23} \\ \mathbf{W}_{31} & \mathbf{W}_{32} & \mathbf{W}_{33} \end{pmatrix}, \quad \underline{\beta} = \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \underline{\beta}_3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{pmatrix}$$

where $\mathbf{G}_{11}$, $\mathbf{W}_{22}$ and $\mathbf{W}_{33}$ are square matrices with k, $(r-k-n_i)$ and n_i rows, respectively. Then $f(\underline{\beta}_{(i)}|H_i, \sigma) = f(\underline{\beta}\,|H_0, \sigma, \underline{\beta}_3 = 0)$ and $\underline{\beta}_{(i)} = (\underline{\beta}_1', \underline{\beta}_2')'$. Since $f(\underline{\beta}\,|H_0, \sigma)$ is a multivariate-normal distribution,

$$\text{(A.1)} \qquad [S(\underline{\beta}_{(i)}|H_i]^{-1} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}$$

and

$$\text{(A.2)} \qquad C(\underline{\beta}_{(i)}|H_i) = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{G}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{W}_{13} \\ \mathbf{W}_{23} \end{pmatrix} \mathbf{b}_3$$

To show (i) we use equation (3.4): the a-CMV point under H_0 is $\{\mathbf{a}', \mathbf{a}'\mathbf{G}_{11}^{-1}(\mathbf{W}_{12}, \mathbf{W}_{13})\}'$ while an a-CMV point under H_i is any point whose first $(r - n_i)$ components equal $(\mathbf{a}', \mathbf{a}'\mathbf{G}_{11}^{-1}\mathbf{W}_{12})'$. Hence the a-CMV point under H_0 is also an a-CMV point under H_i. To show (ii), we first establish that if $\mathbf{x}$ is the CMV-point under H_0, then $S(\overline{y}\,|\,\mathbf{x}, H_0) = S(\overline{y}\,|\,\mathbf{x}, H_i)$ and $C(\overline{y}\,|\,\mathbf{x}, H_0) = C(\overline{y}\,|\,\mathbf{x}, H_i)$. The former clearly holds, since $S(\overline{y}\,|\,\mathbf{x}, H_0)$ and $S(\overline{y}\,|\,\mathbf{x}, H_i)$ both equal $\mathbf{a}'\mathbf{G}_{11}^{-1}\mathbf{a}$, from equation (3.5). For the latter, we have that

$$\begin{aligned} C(\overline{y}\,|\,\mathbf{x}, H_0) &= (\mathbf{b}_1', \mathbf{b}_2', \mathbf{b}_3')\{\mathbf{a}', \mathbf{a}'\mathbf{G}_{11}^{-1}(\mathbf{W}_{12}, \mathbf{W}_{13})\}' \\ &= \mathbf{b}_1'\mathbf{a} + \mathbf{b}_2'\mathbf{W}_{21}\mathbf{G}_{11}^{-1}\mathbf{a} + \mathbf{b}_3'\mathbf{W}_{31}\mathbf{G}_{11}^{-1}\mathbf{a}. \end{aligned}$$

It is straightforward (but tedious) to show that $C(\overline{y}\,|\,\mathbf{x}, H_i)$ also equals this by putting $C(\overline{y}\,|\,\mathbf{x}, H_i) = \{C(\underline{\beta}_{(i)}|H_i)\}'(\mathbf{a}', \mathbf{a}'\mathbf{G}_{11}^{-1}\mathbf{W}_{21}')'$, using equation (A.2), and putting

$$\begin{pmatrix} \mathbf{G}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} [\mathbf{G}_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}]^{-1} & -[\mathbf{G}_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}]^{-1}\mathbf{W}_{12}\mathbf{W}_{22}^{-1} \\ -[\mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{G}_{11}^{-1}\mathbf{W}_{12}]^{-1}\mathbf{W}_{21}\mathbf{G}_{11}^{-1} & [\mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{G}_{11}^{-1}\mathbf{W}_{12}]^{-1} \end{pmatrix}.$$

Both $f(\overline{y}\,|\,\mathbf{x}, H_0)$ and $f(\overline{y}\,|\,\mathbf{x}, H_i)$ are t-distributions on n degrees of freedom, and hence they must be identical since their spreads and centres are equal. This demonstrates (ii). □

Proof of Theorem 2. Let $\mathbf{x}_1$ be the a-CMV point under H_0. Then $f(\overline{y}\,|\,\mathbf{x}_1) = f(\overline{y}\,|\,\mathbf{x}_1, H_i)$ for $i = 0, 1, \ldots, h$. (Theorem 1). Let I be the magnitude of the interquartile ranges of these distributions. Then for all c,

$$\int_c^{c+I} f(\overline{y}\,|\,\mathbf{x}_1, H_i)\,d\overline{y} \le \tfrac{1}{2}.$$

Hence if $\mathbf{x}_2$ is any point whose first k components equal $\mathbf{a}$, the definition of an a-CMV point implies

$$\int_c^{c+I} f(\overline{y}\,|\,\mathbf{x}_2, H_i)\,d\overline{y} \;\leq\; \tfrac{1}{2}.$$

Moreover, the inequality is strict if $\mathbf{x}_2$ differs from $\mathbf{x}_1$ in any component that corresponds to a non-zero β-coefficient under H_i. Hence if $\mathbf{x}_2 \neq \mathbf{x}_1$, then for all c,

$$\sum_{i=0}^{h} \int_c^{c+I} f(\overline{y}\,|\,\mathbf{x}_2, H_i)\,P(H_i)\,d\overline{y} \;<\; \sum_{i=0}^{h} \tfrac{1}{2}P(H_i) \;=\; \tfrac{1}{2}.$$

so the interquartile range of $f(\overline{y}\,|\,\mathbf{x}_2)$ exceeds I. But, by definition, the a-CMV point is the point at which the interquartile range of the predictive distribution of $\overline{y}$ is minimised, subject to the constraint. Hence the a-CMV point is $\mathbf{x}_1$. □

APPENDIX B

Assessing ω and n. To determine the hyperparameters ω and n, the expert is first asked to imagine that two separate experiments will be conducted at the same design point. Let Z_1 be the response in the first experiment minus the response in the second experiment. The expert assesses the median of the unsigned difference $|Z_1|$, his assessment being denoted by k_1. He is asked to imagine that the observed difference was $Z_1 = z_1$ and that two further experiments are to be conducted at a single design point, Z_2 being the difference in the responses these yield. He assesses the median of the magnitude of $Z_2|Z_1 = z_1$, k_2 being his assessment. The value of n is then determined from the equation

$$\frac{k_1}{k_2} = \frac{q_n}{q_{n+1}} \left[\frac{n+1}{(\alpha q_n)^2 + n}\right]^{\frac{1}{2}}$$

where $\alpha = z_1/k_1$ and q_n is the semi-interquartile range of a t-distribution with unit spread and n degrees of freedom. [In the implementation of the method, the computer chooses z_1 so that $\alpha = \frac{1}{2}$ and a table stores the corresponding values of k_1/k_2 for various values of n, thereby simplifying calculations.] After n has been determined, ω is obtained from the equation, $\omega = \frac{1}{2}(k_1/q_n)^2$.

APPENDIX C

DERIVATION OF EQUATIONS (5.5) AND (5.6). To estimate $g_{j,j}$ and $g_{j,i}$ ($j = m+1, \ldots, r$; $i = m+1, \ldots, r$), only assessments at the CMV points $\mathbf{x}_m, \mathbf{x}_{m+1}, \ldots, \mathbf{x}_r$ will be used. The first m components of each of these points equal $1, a_2, a_3, \ldots, a_m$ so the linear model can be restricted to design points that satisfy this constraint. Putting $\alpha = \mathbf{x}'_m\underline{\beta}$, the linear model $E(Y) = \beta_1 + \beta_2 X_2 + \ldots + \beta_r X_r$ becomes

$$E(Y) = \alpha + \beta_{m+1}\xi_{m+1} + \beta_{m+2}\xi_{m+2} + \ldots + \beta_r\xi_r \tag{C.1}$$

where $\xi_j = X_j - x_{m,j}$ for $j = m+1, m+2, \ldots, r$.

The CMV points $\mathbf{x}_m, \mathbf{x}_{m+1}, \ldots, \mathbf{x}_r$ transform to the $(r-m+1)$-dimensional vectors $\eta_m, \eta_{m+1}, \ldots, \eta_r$, where $\eta_m = (1, 0, 0, \ldots, 0)'$ and, for $j = m+1, \ldots, r$, $\eta_j = (1,\ x_{j,m+2} - x_{m,m+1}, \ldots,\ x_{j,j-1} - x_{m,j-1},\ a_j - x_{m,j},\ x_{j,j+1} - x_{m,j+1}, \ldots,\ x_{j,r} - x_{m,r}\ )'$. The MV point for the model in (C.1) is η_m and $\eta_{m+1}, \ldots, \eta_r$ are CMV points. The components of η_j that are constrained are the first component, which is constrained to equal 1, and the $(j-m+1)th$ component, which is constrained to equal $a_j - x_{m,j}$. The spreads of the predictive distributions at these points are given by $S(\overline{y}\,|\,\eta_j, H_0) = S(\overline{y}\,|\,\mathrm{x}_j, H_0)$ for $j = m, \ldots, r$.

Let $\mathbf{G} = [S(\{\alpha, \beta_{m+1}, \ldots, \beta_r\}'\,|H_0)]^{-1}$. Since $(1,0,0,\ldots,0)'$ is the MV point for (C.1), equation (3.4) implies that the off-diagonal elements of the first row and column of $\mathbf{G}$ are zeros. Hence

$$\mathbf{G} = \begin{pmatrix} g_{1,1} & 0 & 0 & \cdots\cdots & 0 \\ 0 & g_{m+1,m+1} & g_{m+1,m+2} & \cdots\cdots & g_{m+1,r} \\ 0 & g_{m+2,m+1} & g_{m+2,m+2} & \cdots\cdots & g_{m+2,r} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & g_{r,m+1} & g_{r,m+2} & \cdots\cdots & g_{r,r} \end{pmatrix}$$

where $g_{11} = [S(\alpha\,|H_0)]$. To estimate this matrix we first note that $\alpha = \mathrm{x}_m'\underline{\beta}$, so

(C.2) $$g_{1,1} = [S(\overline{y}\,|\mathrm{x}_m, H_0)]^{-1}.$$

From equation (3.5), for $j = m+1, m+2, \ldots, r$

$$S(\overline{y}\,|\,\eta_j, H_0) = (1, a_j - x_{m,j}) \begin{pmatrix} g_{1,1} & 0 \\ 0 & g_{j,j} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ a_j - x_{m,j} \end{pmatrix}.$$

Also, from (4.8), $S(\overline{y}\,|\,\eta_j, H_0) - g_{1,1} = S(\overline{y}\,|\mathrm{x}_j, H_0) - S(\overline{y}\,|\,\mathrm{x}_m, H_0) = S(d_j)$, so $g_{j,j} = (a_j - x_{m,j})^2 / S(d_j\,|H_0)$, which is equation (5.5).

From (3.4) and the positions of the CMV-points $\eta_{m+1}, \eta_{m+2}, \ldots, \eta_r$, we have that for $i = m+1, m+2, \ldots, r$; $j = m+1, m+2, \ldots, r$; $i \neq j$,

$$x_{j,i} - x_{m,i} = (0, g_{j,i}) \begin{pmatrix} g_{1,1} & 0 \\ 0 & g_{j,j} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ a_j - x_{m,j} \end{pmatrix}.$$

and equation (5.6) follows. □

PROOF THAT **A** IS NON-SINGULAR. Partition **A** as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are $m \times m$ and $(r-m) \times (r-m)$ matrices, respectively. From the choice of design points (equations (4.1) and (4.3)), $\mathbf{A}_{21} = \mathbf{0}$, so the determinant of **A** equals $|\mathbf{A}_{11}| . |\mathbf{A}_{22}|$. Also, $\mathbf{A}_{11}$ is a diagonal matrix whose diagonal elements are non-zero (they equal 1, $a_2 - x_{1,2}, \ldots, a_m - x_{m-1,m}$), so $|\mathbf{A}_{11}| \neq 0$. Consequently, if $|\mathbf{A}_{22}| \neq 0$, then $|\mathbf{A}|$ is non-zero and hence **A** is non-singular.

During the elicitation procedure it is checked that $\mathbf{V} = S(\{\beta_{m+1}, \ldots, \beta_r\}' \,|H_0)$ is a positive-definite matrix. We relate $\mathbf{A}_{22}$ to **V**. Define the square matrix $\mathbf{Q} = (q_{j,i})$ by

$$q_{j,j} = g_{m+j,m+j}$$

and $$q_{j,i} = g_{m+j,m+j}(x_{m+j,m+i} - x_{m,m+i}) \,/\, (a_{m+j} - x_{m,m+j}),$$

for $j = 1, 2, \ldots, r-m$; $i = 1, 2, \ldots, r-m$; $i \neq j$.

Comparison with (5.7) indicates that $\frac{1}{2}(\mathbf{Q} + \mathbf{Q}') = \mathbf{V}^{-1}$. Since **V** is positive-definite, $0 \neq \psi'\mathbf{V}^{-1}\psi = \frac{1}{2}[\psi'\mathbf{Q}\psi + (\psi'\mathbf{Q}\psi)']$ for any non-zero vector ψ. Consequently, $\mathbf{Q}\psi \neq \mathbf{0}$ for any non-zero vector ψ, so **Q** is non-singular and

$|\mathbf{Q}|$ non-zero. If, for $j = 1, 2, \ldots, r - m$, the jth row of $\mathbf{A}_{22}$ were multiplied by $g_{m+j,m+j}/(a_{m+j} - x_{m,m+j})$, the matrix $\mathbf{Q}$ would be obtained. Hence,

$$|\mathbf{A}_{22}| = |\mathbf{Q}| \prod_{j=1}^{r-m} [(a_{m+j} - x_{m,m+j}) \,/\, g_{m+j,m+j}].$$

Since $|\mathbf{Q}| \neq 0$ and $(a_{m+j} - x_{m,m+j}) \neq 0$ for $j = 1, 2, \ldots, r - m$, we have that $|\mathbf{A}_{22}| \neq 0$. □

Acknowledgements. The authors are grateful to G Thackray, A Baines and other staff of Imperial Chemical Industries at Runcorm who helped to test the elicitation method. This research was supported by the UK Science and Engineering Research Council, by Imperial Chemical Industries and by the US National Science Foundation (grants MCS- 830/335, DMS-8614793, and DMS-8911548).

REFERENCES

ATKINSON, A. C. and FEDOROV, V. V. (1975a). The design of experiments for discriminating between two rival models. *Biometrika* **62** 57–70.

ATKINSON, A. C. and FEDOROV, V. V. (1975b). Optimal design: Experiments for discriminating between several models. *Biometrika* **62** 289–303.

DICKEY, J. M. (1974). Bayesian alternatives to the F test and least-squares estimate in the normal linear model. In *Studies in Bayesian Econometrics and Statistics* (S. E. Fienberg and A. Zellner, eds.) 515–554. North-Holland, Amsterdam.

DICKEY, J. M. (1980). Approximate coherence for regression model inference - With a new analysis of Fisher's Broadbalk Wheatfield example. In *Bayesian Analysis in Econometrics and Statistics* (A. Zellner, ed.) 333–354. North-Holland, Amsterdam.

DICKEY, J. M. and LIENTZ, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Ann. Math. Statist.* **41** 214–226.

GARTHWAITE P. H. (1983). Assessment of Prior Distributions for Normal Linear Models. Ph. D. thesis, Dept. Statistics, Univ. College of Wales, Aberystwyth.

GARTHWAITE, P. H. (1990). Implementation, user guide and computer programme for a method of eliciting opinion about a regression model in variable-selection problems. Technical Report 12, Dept. Math. Sciences, Univ. of Aberdeen.

GARTHWAITE, P. H. and DICKEY, J. M. (1988). Quantifying expert opinion in linear regression problems. *J. Roy. Statist. Soc. Ser. B* **50** 462–474.

GARTHWAITE, P. H. and DICKEY, J. M. (1990). An elicitation method for multiple linear regression models. *J. Behavioral Decision Making.* To appear.

GUNEL, E. and DICKEY, J. M. (1974). Bayes factors for independence in contingency tables. *Biometrika* **61** 545–557.

HILL, B. M. (1974). On coherence, inadmissibility and inference about many parameters in the theory of least squares. In *Studies in Bayesian Econometrics and Statistics* (S. E. Fienberg and A. Zellner, eds.) 555–584. North-Holland, Amsterdam.

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. *J. Amer. Statist. Assoc.* **75** 845–854.

PRESS, S. J. (1972). *Applied Multivariate Analysis.* Holt, Rinehart and Winston, New York.

RAIFFA, H. A. and SCHLAIFER, R. S. (1961). *Applied Statistical Decision Theory.* Graduate School of Business Administration, Harvard University, Boston.

DEPARTMENT OF
MATHEMATICAL SCIENCES
UNIVERSITY OF ABERDEEN
ABERDEEN AB9 2TY
UNITED KINGDOM

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455

TABLE 1
Elicited Points of Constrained Minimum Variance

Point	*Constant*	*Temp 1*	*Time*	*Gas 1*	*Gas 2*	*Temp 2*	*Pres*	*Temp 3*
x_1'	1*	380	8	9	6	320	0	280
x_2'	1*	402.5*	8	9	7	350	0	280
x_3'	1*	402.5*	16*	9	9	320	0	280
x_4'	1*	402.5*	16*	12.5*	10	320	0	320
x_5'	1*	402.5*	16*	12.5*	8*	330	0	280
x_6'	1*	402.5*	16*	12.5*	8*	390*	0	300
x_7'	1*	402.5*	16*	12.5*	8*	350	0.5*	320
x_8'	1*	402.5*	16*	12.5*	8*	350	0	315*

TABLE 2

Median and Quartile Assessments at the Elicited Points of Constrained Minimum Variance

Point	$\overline{y}_{.25}$	$\overline{y}_{.50}$	$\overline{y}_{.75}$	$d_{.25}$	$d_{.50}$	$d_{.75}$
x_1	35	40	44	–	–	–
x_2	33.3*	38	42.7*	−4	−2*	−1
x_3	29.4*	35	40.6*	−8	−3*	−2
x_4	23.6*	30	36.4*	−8	−5*	−2
x_5	20.5*	27	33.5*	−5	−3*	−2
x_6	18.3*	25	31.7*	−4	−2*	−1
x_7	16.3*	23	29.7*	−5	−4*	−2
x_8	19.4*	26*	32.6*	−1.5	−1	−0.5

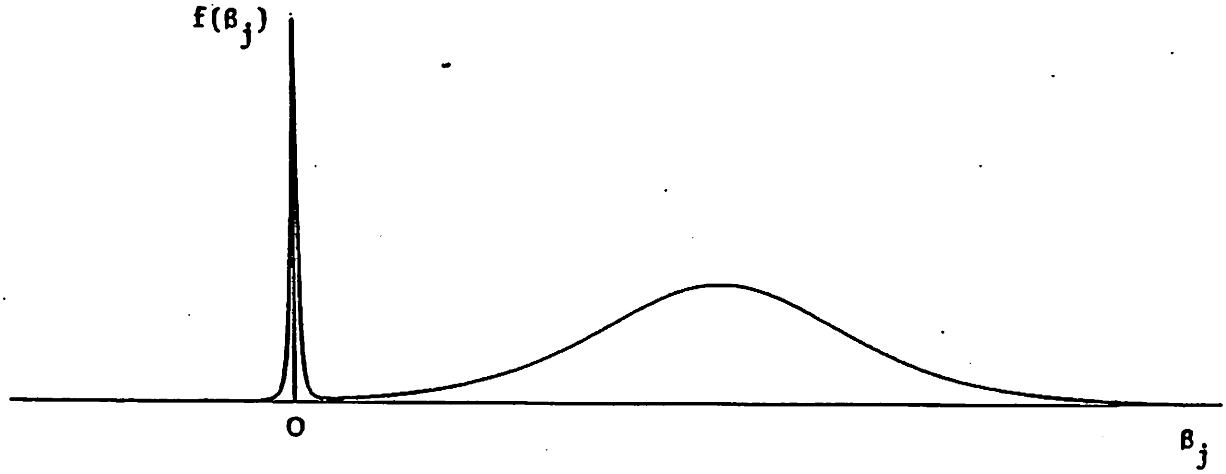

FIG. 1. *Marginal distribution for the coefficient* (β_j) *of a variable that might increase the response, or might have no affect on it.*

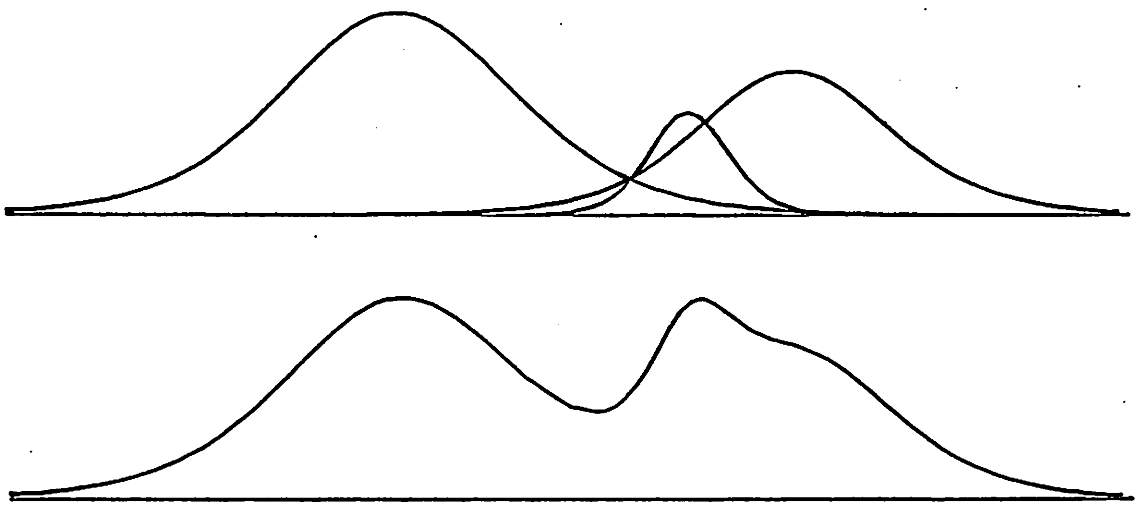

FIG. 2. *Three t-distributions and the mixture distribution they form.*