STUDIES IN
BAYESIAN ECONOMETRICS
AND STATISTICS

Editors

ARNOLD ZELLNER
JOSEPH B. KADANE

Volume 6

NORTH-HOLLAND–AMSTERDAM · NEW YORK · OXFORD

BAYESIAN INFERENCE AND DECISION TECHNIQUES

Essays in Honor of Bruno de Finetti

Edited by

PREM K. GOEL
Ohio State University

ARNOLD ZELLNER
University of Chicago

1986

NORTH-HOLLAND–AMSTERDAM · NEW YORK · OXFORD

39
485

© Elsevier Science Publishers B.V., 1986

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 0 444 87712 6

Published by:

ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

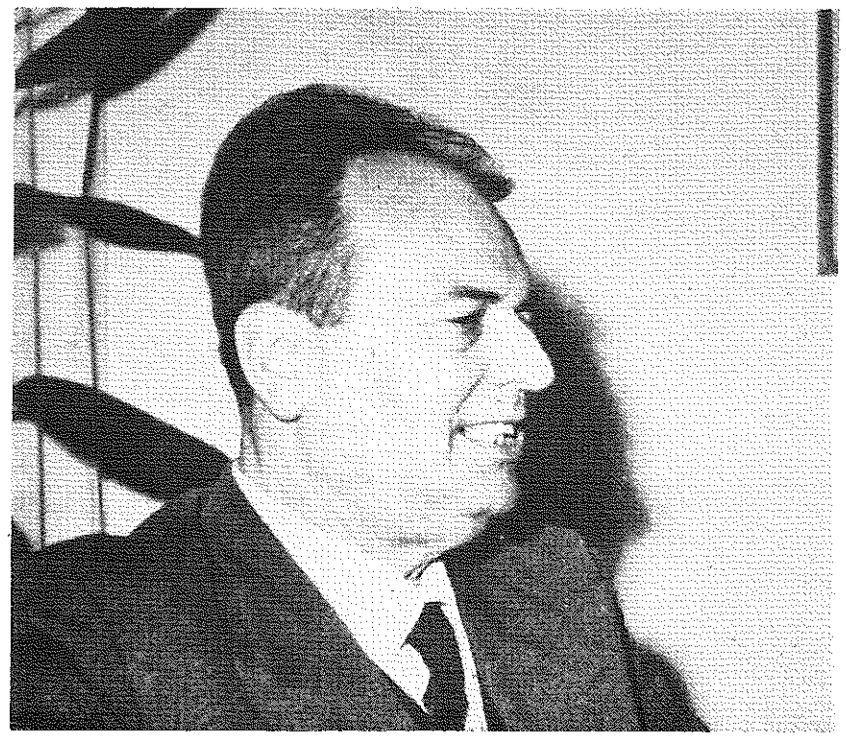
R. U. Groningen
Statistiek
Faculteit Sociale Wetenschappen

PRINTED IN THE NETHERLANDS

INTRODUCTION TO THE SERIES

This Series on Bayesian Econometrics will include monographs reporting significant research on the theory and applications of Bayesian analysis in Econometrics. Research on Bayesian methodology and applications of it has progressed remarkably in the past few decades thus making it appropriate to institute this Series at this time. It is anticipated that the Series will serve as a focal point for bringing significant Bayesian research to the attention of researchers, decision-makers, students and others.

The Editors



Bruno de Finetti 1906–1985

(Photo: courtesy of Massimo De Felice)

Koopman, B.O. (1940b) "The bases of probability", *Bulletin of the American Mathematical Society*, 46, 763–774.

Leamer, E.E. (1978) *Specification Searches* (Wiley, New York).

Leamer, E.E. (1982) "Sets of posterior means with bounded variance priors", *Econometrica*, 5, 725–736.

Marschak, Jacob (1964) "Actual versus consistant decision behavior", *Behavioral Science*, 9, 103–110.

Milgrom, Paul R. and R.J. Weber (1982) "A theory of auctions and competitive bidding", *Econometrica*, 50, 1089–1122.

Raiffa, H. and R. Schlaifer (1961) *Applied Statistical Decision Theory* (Harvard University Press).

Riley, J. and W. Samuelson (1981) "Optimal auctions", *American Economic Review*, 71, 381–392.

Savage, L.J. (1971) "Elicitation of personal probabilities and expectations", *Journal of the American Statistical Association*, 66, 565–614.

Shafer, G. (1976) *A Mathematical Theory of Evidence* (Princeton University Press).

Shafer, G. (1981) "Constructive probability", *Synthese*, 48, 1–60.

Shafer, G. (1982) "Lindley's paradox", *Journal of the American Statistical Association*, 77, 325–351, with comments by D.V. Lindley, Morris H. DeGroot, A.D. Dempster, I.J. Good, Bruce M. Hill, and Robert E. Kass.

Smith, Cedric A.B. (1961) "Consistency in statistical inference and decision", *Journal of the Royal Statistical Society B*, No. 1, 1–25.

Vickrey, W. (1961) "Counterspeculation, auctions, and competitive sealed tenders", *Journal of Finance*, 16, 8–37.

Walley, P. (1981) "Coherent lower (and upper) probabilities", University of Warwick Discussion Paper.

Williams, P.A. (1976) "Indeterminate probabilities", in: M. Przelecki, K. Szaniawski and R. Wojcicki, eds., *Formal Methods in the Methodology of Empirical Science* (Ossolineum and D. Reidel Publishing Co.), pp. 229–246.

Williams, P.A. (1978) "On a new theory of epistemic probability", *British Journal of Philosophy of Science*, 29, 375–387.

Winkler, Robert J.E. (1969) "Scoring roles and the evaluation of probability assessors", *Journal of the American Statistical Association*, 64, 1073–1078.

Wolfenson, M. and T.L. Fine (1982) "Bayes-like decision making with upper and lower probabilities", *Journal of the American Statistical Association*, 77, 80–83.

Chapter 15

ON ASSESSING PRIOR DISTRIBUTIONS AND BAYESIAN REGRESSION ANALYSIS WITH g-PRIOR DISTRIBUTIONS

ARNOLD ZELLNER*

University of Chicago

1. Introduction

In this chapter we consider the normal linear multiple regression model (MRM) and prior information about values of its parameters. As is well known, Bayesian analysis of the MRM when there is little prior information about parameter values has appeared in Jeffreys (1967), Box and Tiao (1973), Zellner (1971), and elsewhere. In cases in which there is prior information about parameter values, prior distributions to represent such information can be and have been employed in analyses of the MRM – see, for example, Raiffa and Schlaifer (1961), Shiller (1973), Leamer (1978), Zellner (1971), Zellner and Geisel (1970), Zellner and Richard (1973), and Zellner and Williams (1973).

Procedures for assessing informative prior distributions for the MRM's parameters have been put forward by Kadane (1980), Kadane et al. (1980), Winkler (1967, 1977), and Zellner (1985). In many instances these procedures will be extremely valuable in enhancing understanding of the particular MRM under consideration and in formulating an appropriate prior distribution for its parameters. In the present chapter another approach for assessing a prior distribution for the parameters of the MRM is put forward. It involves the use of Muth's (1961) rational expectations hypothesis and leads to relatively simple prior distributions that are called "g-priors." It will be seen that g-priors are a special form of a

*This is a revised version of an earlier paper presented to the Econometrics and Statistics Colloquium, University of Chicago on 27 February 1980 and at the Econometric Society Meeting, Denver, Colorado, September 1980. Research financed by NSF Grant SES 7913414 and by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago.

Bayesian Inference and Decision Techniques, edited by P. Goel and A. Zellner

© Elsevier Science Publishers B.V., 1986

natural conjugate prior distribution for the MRM's parameters and can be viewed as reference informative prior distributions.

The plan of the paper is as follows. In section 2 a g-prior distribution is assessed and compared with other prior distributions for the MRM's parameters. The properties of posterior distributions based on a particular g-prior are discussed in section 3. The risk function and average risk of the posterior mean of the regression coefficient vector relative to a quadratic loss function and a g-prior are derived and compared with those of the usual least squares estimator in section 4. In section 5 selected posterior odds ratios involving g-prior distributions are presented and analyzed. Finally, some concluding remarks are presented in section 6.

2. Prior distributions for the multiple regression model (MRM)

The standard MRM can be written as

$$y = X\boldsymbol{\beta} + \boldsymbol{u}, \tag{2.1}$$

where $\boldsymbol{y}$ is an $n\times 1$ vector of observations on the dependent variable, X is an $n\times k$ given matrix, assumed of rank k, $\boldsymbol{\beta}$ is a $k\times 1$ vector of regression parameters with unknown values, and $\boldsymbol{u}$ is an $n\times 1$ vector of disturbance or error terms. It is assumed that the elements of $\boldsymbol{u}$ have been independently drawn from a normal distribution with zero mean and finite unknown variance σ^2. The case of normal error terms will be treated herein, but it should be appreciated that other distributional assumptions can also be utilized – see, for example, Zellner (1976). The likelihood function for the MRM is given by

$$\begin{aligned} l(\boldsymbol{\beta},\sigma|y,X) &\propto \sigma^{-n}\exp\{-(y-X\boldsymbol{\beta})'(y-X\boldsymbol{\beta})/2\sigma^2\} \\ &\propto \sigma^{-n}\exp\{-[\nu s^2+(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})]/2\sigma^2\}, \end{aligned} \tag{2.2}$$

where $\hat{\boldsymbol{\beta}}=(X'X)^{-1}X'y$, $\nu s^2=(y-X\hat{\boldsymbol{\beta}})'(y-X\hat{\boldsymbol{\beta}})$, and $\nu=n-k$.

We now turn to discuss several prior distributions for $\boldsymbol{\beta}$ and σ. First, in cases in which little is known about the values of $\boldsymbol{\beta}$ and σ, Jeffreys (1967, pp. 147ff.) recommends the following "diffuse" or "vague" prior distribution, which is widely employed – see, for example, Lindley (1965), DeGroot (1970), Zellner (1971), and Box and Tiao (1973):

$$p(\boldsymbol{\beta},\sigma)\propto 1/\sigma, \tag{2.3}$$

that is, the elements of $\boldsymbol{\beta}$ and $\log\sigma$ are assumed independently and uniformly distributed.[1] The prior in (2.3) can be obtained from Jeffreys' (1967, p. 180) general rule, $|\mathrm{Inf}|^{1/2}$ for generating invariant diffuse prior distributions, where

[1] Jeffreys (1967, p. 122) explains that the ranges of the parameters in (2.2) must be finite in order to maintain his Convention 3. However, the ranges can be very large so that using $-\infty<\beta_i<\infty$, $i=1,2,\ldots,k$, and $0<\sigma<\infty$ in integrations will usually provide accurate results.

Inf is the Fisher information matrix. In the present instance the rule is applied separately to $\boldsymbol{\beta}$ and to σ to yield (2.3). Also, the diffuse prior in (2.3) is produced by the maximal data information procedure described in Zellner (1971, ch. 2, and 1977).

If we consider combining the diffuse prior in (2.3) with the likelihood function in (2.2) *before* we take the data, then we know that the posterior distribution for $\boldsymbol{\beta}$ given σ will be normal with mean $\hat{\boldsymbol{\beta}}=(X'X)^{-1}X'y$ and covariance matrix $(X'X)^{-1}\sigma^2$. Thus, we know beforehand that the elements of $\boldsymbol{\beta}$ will be correlated in the posterior distribution in the usual case in which $X'X$ is not a diagonal matrix and know the form of the covariance matrix.[2]

A second prior distribution for $\boldsymbol{\beta}$ and σ is the natural conjugate prior distribution of Raiffa and Schlaifer (1961). It is given by

$$p(\boldsymbol{\beta},\sigma)=p_{\mathrm{N}}(\boldsymbol{\beta}|\sigma)\,p_{\mathrm{IG}}(\sigma), \tag{2.4a}$$

with

$$p_{\mathrm{N}}(\boldsymbol{\beta}|\sigma)\propto\sigma^{-k}\exp\{-(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}})'A(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}})/2\sigma^2\} \tag{2.4b}$$

and

$$p_{\mathrm{IG}}(\sigma)\propto\sigma^{-(\nu_0+1)}\exp\{-\nu_0 s_0^2/2\sigma^2\}, \tag{2.4c}$$

where $p_{\mathrm{N}}(\boldsymbol{\beta}|\sigma)$ is a multivariate normal pdf for $\boldsymbol{\beta}$ given σ with prior mean $\bar{\boldsymbol{\beta}}$ and prior covariance matrix $A^{-1}\sigma^2$ (the matrix A is assumed to be positive definite symmetric) and $p_{\mathrm{IG}}(\sigma)$ in (2.4c) is in the inverted gamma form with positive parameters ν_0 and s_0^2.

If (2.4) is combined with the likelihood function in (2.2), the posterior pdf for $\boldsymbol{\beta}$ and σ is of well-known form. In particular, the posterior mean for $\boldsymbol{\beta}$, denoted by $\bar{\bar{\boldsymbol{\beta}}}$ is given by

$$\begin{aligned} \bar{\bar{\boldsymbol{\beta}}} &= (A+X'X)^{-1}(A\bar{\boldsymbol{\beta}}+X'X\hat{\boldsymbol{\beta}}) \\ &= \bar{\boldsymbol{\beta}}+[I-(A+X'X)^{-1}A](\hat{\boldsymbol{\beta}}-\bar{\boldsymbol{\beta}}), \end{aligned} \tag{2.5}$$

where $\hat{\boldsymbol{\beta}}=(X'X)^{-1}X'y$ and $\bar{\boldsymbol{\beta}}$ is the prior mean of $\boldsymbol{\beta}$.[3] Thus, $\bar{\bar{\boldsymbol{\beta}}}$ can be viewed as a "shrinkage" estimate. Also, the relation of $\bar{\bar{\boldsymbol{\beta}}}$ to "ridge-regression" estimates is well known. Furthermore, it is known that $\bar{\bar{\boldsymbol{\beta}}}$, viewed as an estimator, is optimal relative to quadratic loss functions in the sense that it minimizes average risk and thus is admissible – see Giles and Rayner (1979) for an analysis of the sampling properties of $\bar{\bar{\boldsymbol{\beta}}}$ relative to those of $\hat{\boldsymbol{\beta}}$. In addition, in large samples, the posterior pdf for $\boldsymbol{\beta}$ is normal, centered at $\hat{\boldsymbol{\beta}}$, the maximum likelihood estimate, and has a covariance matrix proportional to $(X'X)^{-1}$.

[2] If the elements of X can be chosen by an investigator, he can control the form of the posterior covariance matrix for $\boldsymbol{\beta}$.

[3] Using the natural conjugate prior distribution in (2.4), the posterior covariance matrix for $\boldsymbol{\beta}$ is proportional to $(A+X'X)^{-1}$. Since the elements of A are of order 1, while those of $X'X$ are usually of order n, the form of $X'X$ will have a pronounced influence on the form of the posterior covariance matrix for $\boldsymbol{\beta}$. See Zellner and Williams (1973) for an application illustrating this point.

While procedures for assessing the natural conjugate prior in (2.4) have been developed – see Winkler (1967, 1977), Kadane et al. (1980) and Zellner (1985) – for many regression models it is difficult to evaluate the prior covariances for the elements of β. For such situations and others, we propose the following procedure for assessing a reference informative prior distribution. Before observing y, consider a "conceptual" or "imaginary" sample, y_0, an $n\times 1$ vector, assumed generated by

$$y_0 = X\beta + u_0, \tag{2.6}$$

where u_0 is assumed drawn from $N(0,\sigma_0^2 I_n)$. In (2.6), X is the same design matrix appearing in (2.2) since it is assumed that the MRM is appropriate for the given values of the independent variables in X. However, since y_0 is a conceptual sample, we allow σ_0^2 to be different from σ^2 and assume $\sigma^2 = g\sigma_0^2$, $0<g<\infty$, where g will initially be assumed given. Then $\sqrt{g}\,u_0 = \varepsilon$ is $N(0,\sigma^2 I_n)$ and with a diffuse prior for β and σ, $p(\beta,\sigma)\propto 1/\sigma$, the "posterior" pdf based on the conceptual sample in (2.6) is:

$$\begin{aligned} p(\beta,\sigma|D_0) &\propto \sigma^{-(n+1)}\exp\{-g(y_0-X\beta)'(y_0-X\beta)/2\sigma^2\} \\ &\propto \sigma^{-(n+1)}\exp\{-g[\nu s_0^2 + (\beta-\hat\beta_0)'X'X(\beta-\hat\beta_0)]/2\sigma^2\}, \end{aligned} \tag{2.7}$$

where D_0 denotes the sample information in (2.6) and the diffuse prior assumptions, $\nu = n-k$, $\hat\beta_0 = (X'X)^{-1}X'y_0$ and $\nu s_0^2 = (y_0 - X\hat\beta_0)'(y_0-X\hat\beta_0)$.

Now suppose that we have given, anticipated values for β and σ^2, denoted by β_a and σ_a^2, respectively, in addition to (2.6). Application of Muth's (1961) rational expectations hypothesis leads us to take

$$\beta_a = E(\beta|D_0) = \hat\beta_0 \tag{2.8}$$

and

$$\sigma_a^2 = E(\sigma^2|D_0) = \nu g s_0^2/\nu - 2, \tag{2.9}$$

where $E(\beta|D_0)$ and $E(\sigma^2|D_0)$ are posterior means derived from (2.7). From (2.9), $gs_0^2 = (\nu-2)\sigma_a^2/\nu \equiv \bar\sigma_a^2$ and the joint reference informative g-prior distribution is:

$$\begin{aligned} p(\beta,\sigma|\theta_0) &\propto \sigma^{-(\nu+1)}\exp\{-\nu\bar\sigma_a^2/2\sigma^2\} \\ &\quad\times \sigma^{-k}\exp\{-g(\beta-\beta_a)'X'X(\beta-\beta_a)/2\sigma^2\}, \end{aligned} \tag{2.10}$$

where $\theta_0' = (\beta_a', \bar\sigma_a, g, \nu)$. The marginal prior pdfs for σ and β are:

$$p(\sigma|\bar\sigma_a,\nu) \propto \sigma^{-(\nu+1)}\exp\{-\nu\bar\sigma_a^2/2\sigma^2\} \tag{2.11}$$

and

$$p(\beta|\beta_a,g,\nu) \propto \{\nu\bar\sigma_a^2 + g(\beta-\beta_a)'X'X(\beta-\beta_a)\}^{-(\nu+k)/2}. \tag{2.12}$$

It is seen that the joint prior in (2.10) is in the inverted-gamma-normal form with the marginal prior for σ in (2.11) inverted gamma and that for β in (2.12) multivariate Student-t. From (2.12), $E\beta = \beta_a$ and $E(\beta-\beta_a)(\beta-\beta_a)' = (X'X)^{-1}\nu\bar\sigma_a^2/g(\nu-2) = (X'X)^{-1}\sigma_a^2/g$ and from (2.11), $E\sigma^2 = \nu\bar\sigma_a^2/(\nu-2) = \sigma_a^2$.

Thus, given β_a, σ_a^2, (2.6) and use of rational expectations, the complete prior for β and σ in (2.10) has been assessed and is in the natural conjugate form (2.4) with $\bar\beta = \beta_a$, $A = gX'X$, $\nu_0 s_0^2 = \nu\bar\sigma_a^2$ and $\nu_0 = \nu$. Furthermore, note that:

(1) When g's value is unknown, a prior pdf for g, uninformative or informative, say, in the gamma form, can be introduced and g can be integrated out.

(2) One can rewrite (2.6) as $y_0 = X_0\beta + u_0$, with y_0 and u_0 each of dimension $n_0\times 1$ and X_0 of dimension $n_0\times k$ and proceed to derive a prior pdf for the regression parameters. This involves the assumption that the form of the regression model is the same for both design matrices X_0 and X.

(3) It is possible to have $g = g(n)$, e.g. $g \propto 1/n$ or $g\propto \log n/n$ assumptions that control the dependence of the prior precision on n.

Having indicated how to assess g-prior distributions, we now turn to discuss properties of posterior distributions based on them.

3. Posterior distribution for β and σ based on a g-prior distribution

In this section we employ a particular g-prior pdf, namely $p(\beta,\sigma) = h(\sigma)f(\beta|\sigma,g)$, with $h(\sigma)\propto 1/\sigma$ and $f(\beta|\sigma,g)\propto\sigma^{-k}\exp\{-g(\beta-\bar\beta)'X'X(\beta-\bar\beta)$, to obtain the following joint posterior pdf for β and σ:

$$\begin{aligned} p(\beta,\sigma|D) &\propto p_g(\beta,\sigma)l(\beta,\sigma|y) \\ &\propto \sigma^{-(n+k+1)}\exp\{-[(y-X\beta)'(y-X\beta) \\ &\qquad + g(\beta-\bar\beta)'X'X(\beta-\bar\beta)]/2\sigma^2\}, \end{aligned} \tag{3.1}$$

where D denotes the data and prior information. If we let $w' = (y' \vdots g^{1/2}\bar\beta'X')$ and $W' = (X' \vdots g^{1/2}X')$, then the terms in square brackets in the exponential can be expressed as

$$(w-W\beta)'(w-W\beta) = (w-W\bar{\bar\beta})'(w-W\bar{\bar\beta}) + (\beta-\bar{\bar\beta})'W'W(\beta-\bar{\bar\beta}),$$

where $\bar{\bar\beta} = (W'W)^{-1}W'w$. Thus, (3.1) can be expressed conveniently as

$$\begin{aligned} p(\beta,\sigma|D) &\propto \sigma^{-(n+k+1)}\exp\{-[(w-W\bar{\bar\beta})'(w-W\bar{\bar\beta}) \\ &\qquad + (\beta-\bar{\bar\beta})'W'W(\beta-\bar{\bar\beta})]/2\sigma^2\}, \end{aligned} \tag{3.2}$$

where

$$\bar{\bar\beta} = (W'W)^{-1}W'w = (\hat\beta + g\bar\beta)/(1+g), \tag{3.3}$$

with $\hat\beta = (X'X)^{-1}X'y$.

The quantity $\bar{\bar\beta}$ in (3.3) is the mean of the posterior pdf. It is seen to be a simple average of $\hat\beta$, the least squares quantity, and $\bar\beta$, the prior mean vector. Furthermore, the covariance matrix of the conditional normal posterior pdf for β given σ, denoted by $V(\boldsymbol{\beta}|\sigma, D)$, is

$$V(\boldsymbol{\beta}|\sigma, D) = (W'W)^{-1}\sigma^2$$
$$= (X'X)^{-1}\sigma^2/(1+g), \tag{3.4}$$

a very simple result which shows how the size of g affects the spread of the conditional posterior pdf for $\boldsymbol{\beta}$ given σ.

The marginal posterior pdf for $\boldsymbol{\beta}$, obtained from (3.2) by integrating with respect to σ, is

$$p(\boldsymbol{\beta}|D) \propto \{(\boldsymbol{w} - W\bar{\bar{\boldsymbol{\beta}}})'(\boldsymbol{w} - W\bar{\bar{\boldsymbol{\beta}}}) + (\boldsymbol{\beta} - \bar{\bar{\boldsymbol{\beta}}})'W'W(\boldsymbol{\beta} - \bar{\bar{\boldsymbol{\beta}}})\}^{-(n+k)/2}, \tag{3.5}$$

which is in the multivariate Student-t form with mean vector $\bar{\bar{\boldsymbol{\beta}}}$ shown in (3.3) and covariance matrix $V(\boldsymbol{\beta}|D) = (W'W)^{-1}a^2 = (X'X)^{-1}a^2/(1+g)$, where $(n-2)a^2 \equiv (\boldsymbol{w} - W\bar{\bar{\boldsymbol{\beta}}})'(\boldsymbol{w} - W\bar{\bar{\boldsymbol{\beta}}}) = (\boldsymbol{y} - X\bar{\bar{\boldsymbol{\beta}}})'(\boldsymbol{y} - X\bar{\bar{\boldsymbol{\beta}}}) + g(\bar{\bar{\boldsymbol{\beta}}} - \bar{\boldsymbol{\beta}})'X'X(\bar{\bar{\boldsymbol{\beta}}} - \bar{\boldsymbol{\beta}})$. Furthermore, the marginal posterior pdf for σ, obtained by integrating (3.2) with respect to the elements of $\boldsymbol{\beta}$, is

$$p(\sigma|D) \propto \sigma^{-(n+1)}\exp\{-(\boldsymbol{w} - W\bar{\bar{\boldsymbol{\beta}}})'(\boldsymbol{w} - W\bar{\bar{\boldsymbol{\beta}}})/2\sigma^2\}, \tag{3.6}$$

a posterior pdf in the inverted gamma form – see, for example, Raiffa and Schlaifer (1961) and Zellner (1971) for a review of its properties. For (3.6), the posterior mean of σ^2 is given by $E(\sigma^2|D) = a^2$, where a^2 has been defined above. Note that the value of a^2 depends on g.

Thus, from what has been presented, it is direct to use a g-prior distribution to obtain simple posterior distributions, much the same as in the case of the use of a natural conjugate prior distribution. The posterior mean of the coefficient vector shown in (3.3) is a simple average of the prior mean vector $\bar{\boldsymbol{\beta}}$ and of the least squares estimate $\hat{\boldsymbol{\beta}}$ with the parameter g involved in the weights. If g is large, the posterior mean is close to the prior mean $\bar{\boldsymbol{\beta}}$, while if g is small, it will be close to the least squares estimate. Also, g enters simply in expressions for the dispersion of the posterior distribution of $\boldsymbol{\beta}$ and in the posterior pdf for σ.

We now turn to consider some sampling properties of the posterior mean in (3.3).

4. Sampling properties of the posterior mean

It is well known that the mean of the posterior distribution, $\bar{\bar{\boldsymbol{\beta}}}$ in (3.3), is an optimal point estimate relative to quadratic loss in the senses of minimizing posterior expected loss and minimizing average (or Bayes) risk for any given value of σ. Below we derive some sampling properties of $\bar{\bar{\boldsymbol{\beta}}}$ including the risk function for $\bar{\bar{\boldsymbol{\beta}}}$ relative to a quadratic loss function and compare them with those of the least squares estimator, results similar to those of Giles and Rayner (1979).

First, the sampling mean of $\bar{\bar{\boldsymbol{\beta}}}$ is given by

$$E\bar{\bar{\boldsymbol{\beta}}} = (\boldsymbol{\beta}_0 + g\bar{\boldsymbol{\beta}})/(1+g), \tag{4.1}$$

Table 15.1. Values of $(1+g^2\lambda)/(1+g)^2$ for various values of λ and g.

λ	$g=0.10$	$g=0.25$	$g=0.50$	$g=0.75$	$g=1.00$
0	0.826	0.640	0.444	0.327	0.250
1	0.835	0.680	0.556	0.510	0.500
2	0.843	0.720	0.667	0.694	0.750
3	0.851	0.760	0.778	0.878	1.000
4	0.860	0.800	0.889	1.061	1.250
5	0.868	0.840	1.000	1.245	1.500
6	0.876	0.880	1.111	1.429	1.750
7	0.884	0.920	1.222	1.612	2.000
8	0.893	0.960	1.333	1.796	2.250
9	0.900	1.000	1.444	1.980	2.500

where $\boldsymbol{\beta}_0$ is the true unknown value of the regression coefficient vector. Thus, the bias of $\bar{\bar{\boldsymbol{\beta}}}$ is

$$\boldsymbol{B}(\bar{\bar{\boldsymbol{\beta}}}) = E\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta}_0 = g(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)/(1+g). \tag{4.2}$$

Clearly, the bias approaches zero as $g \to 0$ and/or $\bar{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$.

Second, the second moment matrix of $\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta}_0$ is

$$E(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta}_0)(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta}_0)' = V(\hat{\boldsymbol{\beta}})/(1+g)^2 + \boldsymbol{B}(\bar{\bar{\boldsymbol{\beta}}})\boldsymbol{B}(\bar{\bar{\boldsymbol{\beta}}})', \tag{4.3}$$

with $V(\hat{\boldsymbol{\beta}}) = (X'X)^{-1}\sigma^2$, the covariance matrix of the least squares estimator.

Third, the variance–covariance matrix of $\bar{\bar{\boldsymbol{\beta}}}$ is

$$E(\bar{\bar{\boldsymbol{\beta}}} - E\bar{\bar{\boldsymbol{\beta}}})(\bar{\bar{\boldsymbol{\beta}}} - E\bar{\bar{\boldsymbol{\beta}}})' = V(\hat{\boldsymbol{\beta}})/(1+g)^2. \tag{4.4}$$

Fourth, the risk function for $\bar{\bar{\boldsymbol{\beta}}}$ relative to the loss function $L(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}) = (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is conveniently evaluated by noting that $E(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta})'(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta}) = \operatorname{tr}\{E(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta})(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta})'\}$. It follows from (4.3) that

$$E(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta})'(\bar{\bar{\boldsymbol{\beta}}} - \boldsymbol{\beta}) = c\sigma^2(1 + g^2\boldsymbol{\delta}'\boldsymbol{\delta}/c\sigma^2)/(1+g)^2, \tag{4.5}$$

where $c \equiv \operatorname{tr}(X'X)^{-1}$ and $\boldsymbol{\delta} \equiv \bar{\boldsymbol{\beta}} - \boldsymbol{\beta}$.

For comparison with (4.5), the risk function for the least squares estimator is $c\sigma^2$. Thus, it is seen that the risk associated with $\bar{\bar{\boldsymbol{\beta}}}$ will be smaller than that associated with $\hat{\boldsymbol{\beta}}$ if

$$(1 + g^2\boldsymbol{\delta}'\boldsymbol{\delta}/c\sigma^2)/(1+g)^2 < 1. \tag{4.6}$$

Shown in table 15.1 are values of the quantity $(1+g^2\lambda)/(1+g)^2$, where $\lambda \equiv \boldsymbol{\delta}'\boldsymbol{\delta}/c\sigma^2$ for different values of g and λ.[4] From the tabled values, it is seen that $\bar{\bar{\boldsymbol{\beta}}}$ has much lower risk than $\hat{\boldsymbol{\beta}}$ when $\lambda = 0$ and when g is large. However, as λ grows in value, the risk of $\bar{\bar{\boldsymbol{\beta}}}$ grows. It grows more rapidly with increasing λ the

[4]Given a g-prior for $\boldsymbol{\beta}$ and σ, the prior pdf for λ that is in the form of a non-central χ^2 pdf can be employed to make prior inferences about the value of λ.

larger is the value of g. When $\lambda = 1+2/g$, the risks associated with $\bar{\bar{\beta}}$ and $\hat{\beta}$ are equal.[5]

Now, using the g-prior distribution for β given σ, shown above (3.1), it is possible to compute the average of the risk function, (4.5) for $\bar{\bar{\beta}}$. The resulting expression for average (or Bayes) risk of $\bar{\bar{\beta}}$ given σ, denoted by $\mathrm{AR}(\bar{\bar{\beta}}|\sigma)$ is

$$\begin{aligned}\mathrm{AR}(\bar{\bar{\beta}}|\sigma) &= c\sigma^2(1+g^2E\delta'\delta/c\sigma^2)/(1+g)^2 \\ &= c\sigma^2/(1+g), \qquad (4.7)\end{aligned}$$

since $E\delta'\delta = E(\beta-\bar{\beta})'(\beta-\bar{\beta}) = c\sigma^2/g$, where $c = \mathrm{tr}(X'X)^{-1}$. Given σ, the average risk of the least squares estimator is

$$\mathrm{AR}(\hat{\beta}|\sigma) = c\sigma^2. \qquad (4.8)$$

On comparing (4.7) and (4.8), it is seen that there can be a considerable reduction in average risk associated with the use of $\bar{\bar{\beta}}$ relative to use of $\hat{\beta}$.

Finally, we note that $\bar{\bar{\beta}}$ in (3.3) can be generated as the solution to the following constrained least squares problem. Minimize $(y-X\beta)'(y-X\beta)$ with respect to β subject to $(\beta-\bar{\beta})'X'X(\beta-\bar{\beta}) \le r^2$, with $\bar{\beta}$ and r^2 given. Write $(y-X\beta)'(y-X\beta)+g[(\beta-\bar{\beta})'X'X(\beta-\bar{\beta})-r^2]$, where g is interpreted as a Lagrange multiplier. The value of β which sets the derivative of the Lagrangian function with respect to β equal to zero is $\bar{\bar{\beta}} = \bar{\beta}+(\hat{\beta}-\bar{\beta})/(1+g) = (\hat{\beta}+g\bar{\beta})/(1+g)$, with $\hat{\beta} = (X'X)^{-1}X'y$.

5. Posterior odds ratios and g-prior distributions

Herein, we present posterior odds ratios for selected pairs of hypotheses about the values of the regression coefficient vector β.

We first consider the following two hypotheses: H_1: $\beta = \bar{\beta}$ and H_2: $\beta \in R^k$. Under H_1 and H_2 we assume $h(\sigma) \propto 1/\sigma$, and under H_2 we use the following g-prior:

$$p(\beta|\sigma,\bar{\beta}) = (2\pi g/\sigma^2)^{k/2}|X'X|^{1/2}\exp\{-g(\beta-\bar{\beta})'X'X(\beta-\bar{\beta})/2\sigma^2\},$$

where $\bar{\beta}$ is the prior mean vector.[6] With a prior odds ratio of 1:1, the posterior

[5] In terms of prediction of a future value of the dependent variable, $y_f = x_f'\beta_0 + u_f$, with x_f' given, the least squares forecast error is $e_{LS} = y_f - \hat{y}_f = x_f'(\hat{\beta}-\beta_0) - u_f$, while the forecast error associated with the use of $\bar{\bar{\beta}}$ as a point estimator for β is $e_g = x_f'(\bar{\bar{\beta}}-\beta_0) - u_f$. Then $Ee_{LS}^2 = x_f'V(\hat{\beta})x_f + \sigma^2$ and $Ee_g^2 = x_f'E(\bar{\bar{\beta}}-\beta_0)(\bar{\bar{\beta}}-\beta_0)'x_f + \sigma^2$, with $E(\bar{\bar{\beta}}-\beta_0)(\bar{\bar{\beta}}-\beta_0)'$ as shown in (4.3). Risk comparisons for the two predictors can be made for any given value of x_f.

[6] Zellner and Siow (1979) and Mayer (1980) have employed prior pdfs depending on $X'X$ in deriving posterior odds ratios for hypotheses about regression coefficients' values but have not employed g-priors that are utilized in the present chapter.

odds ratio, K_{12}, is given by

$$\begin{aligned}K_{12} &= \left(\frac{1+g}{g}\right)^{k/2}\left[\frac{(w-W\bar{\bar{\beta}})'(w-W\bar{\bar{\beta}})}{(y-X\hat{\beta})'(y-X\hat{\beta})}\right]^{n/2} \\ &= \left(\frac{1+g}{g}\right)^{k/2}\left[\frac{1+\left(\frac{g}{1+g}\right)\frac{k}{\nu}F}{1+\frac{k}{\nu}F}\right]^{n/2}, \qquad (5.1)\end{aligned}$$

with $\nu = n-k$, $F = (\hat{\beta}-\bar{\beta})'X'X(\hat{\beta}-\bar{\beta})/ks^2$, $\hat{\beta} = (X'X)^{-1}X'y$, $\nu s^2 = (y-X\hat{\beta})'(y-X\hat{\beta})$, $\bar{\bar{\beta}} = (\hat{\beta}+g\bar{\beta})/(1+g)$, $w' = (y' \vdots g^{1/2}\bar{\beta}'X')$ and $W' = (X' \vdots g^{1/2}X')$.

In (5.1) F is the usual F-statistic for testing $\beta = \bar{\beta}$. As F grows in value, $K_{12} \to [g/(1+g)]^{n-k/2}$ which will be small for moderate to large n thus providing support for the less restrictive hypothesis, H_2. Also, as $g \to \infty$, $K_{12} \to 1$, which is reasonable since H_1 and H_2 become equivalent as $g \to \infty$.

As pointed out in Zellner and Vandaele (1975), the following Bayesian "pre-test" estimator is optimal relative to quadratic loss:

$$\begin{aligned}\tilde{\beta}^* &= p_1\bar{\beta}+(1-p_1)\bar{\bar{\beta}} \\ &= \bar{\beta}+(\bar{\bar{\beta}}-\bar{\beta})/(1+K_{12}), \qquad (5.2)\end{aligned}$$

where p_1 and $1-p_1$ are posterior probabilities associated with H_1 and H_2, respectively, $K_{12} = p_1/(1-p_1)$ and $\bar{\bar{\beta}} = (\hat{\beta}+g\bar{\beta})/(1+g)$ is the posterior mean under H_2.

We now present a posterior odds ratio for the following two hypotheses: H_A: $\beta = \beta_\mathrm{A}$ and H_B: $\beta \in R^k$, using the same priors as above, where β_A is an arbitrary value for β, not necessarily equal to the prior mean as in H_1 above. With prior odds 1:1, the posterior odds ratio, K_{AB}, for H_A relative to H_B is

$$\begin{aligned}K_{\mathrm{AB}} &= \left(\frac{1+g}{g}\right)^{k/2}\left[\frac{(w-W\bar{\bar{\beta}})'(w-W\bar{\bar{\beta}})}{(y-X\beta_\mathrm{A})'(y-X\beta_\mathrm{A})}\right]^{n/2} \\ &= \left(\frac{1+g}{g}\right)^{k/2}\left[\frac{1+\left(\frac{g}{1+g}\right)\frac{k}{\nu}F_2}{1+\frac{k}{\nu}F_1}\right]^{n/2}, \qquad (5.3)\end{aligned}$$

where, in addition to the quantities defined above,

$$F_1 = (\hat{\beta}-\beta_\mathrm{A})'X'X(\hat{\beta}-\beta_\mathrm{A})/ks^2$$

and

$$F_2 = (\hat{\beta}-\bar{\beta})'X'X(\hat{\beta}-\bar{\beta})/ks^2.$$

It is seen that F_1 and F_2 are the usual F-statistics for the hypotheses $\beta = \beta_\mathrm{A}$ and $\beta = \bar{\beta}$, respectively. As F_2 grows in value, K_{AB} grows in value lending support to

H_A. Similarly, as F_1 grows in value, K_{AB}'s value declines providing support for H_B. As $g \to \infty$,

$$K_{AB} \to \left[\left(1+\frac{k}{\nu}F_2\right)\Big/\left(1+\frac{k}{\nu}F_1\right)\right]^{n/2},$$

which is the odds ratio for the two simple hypotheses $\boldsymbol{\beta} = \boldsymbol{\beta}_A$ and $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$.

As above, it is possible to "average" over the two hypotheses to obtain an estimator that is optimal relative to quadratic loss, namely $\boldsymbol{\beta}^* = \boldsymbol{\beta}_A + (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_A)/(1 + K_{AB})$.

6. Concluding remarks

In this chapter procedures for analyzing the multiple regression model when prior information about its parameters' values is available have been reviewed. A general procedure for assessing prior pdfs for regression coefficients was presented that produced a prior pdf for $\boldsymbol{\beta}$ with a precision matrix proportional to $X'X$ that is termed a g-prior pdf. Use of the g-prior pdf leads to a posterior mean for the regression coefficient vector that is a simple average of the least squares estimate and the prior mean vector. Other properties of the posterior are relatively simple. Also, sampling properties of the posterior mean, derived from a g-prior pdf have been established and compared with those of the least squares estimator including their risk functions and Bayes risk. In addition, posterior odds ratios involving g-prior pdfs have been derived. From these results it appears that g-prior pdfs will be useful in analyzing the multiple regression model in a number of applied problems.

References

Box, G.E.P. and G.C. Tiao (1973) *Bayesian Inference in Statistical Analysis* (Addison-Wesley Publishing Co., Reading, Mass.).

DeGroot, M.H. (1970) *Optimal Statistical Decisions* (McGraw-Hill Book Co., New York.).

Giles, D.E.A. and A.C. Rayner (1979) "The mean squared errors of the maximum likelihood and natural conjugate Bayes regression estimators", *Journal of Econometrics*, 11, 319–334.

Jeffreys, H. (1967) *Theory of Probability* (Oxford University Press, London).

Kadane, J.B. (1980), "Predictive and structural methods for eliciting prior distributions", in: A. Zellner, ed., *Bayesian Analysis in Econometrics and Statistics in Honor of Harold Jeffreys* (North-Holland Publishing Co., Amsterdam), pp. 89–93.

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

Leamer, E.E. (1978) *Specification Searches: Ad Hoc Inference with Nonexperimental Data* (John Wiley & Sons, Inc., New York).

Lindley, D.V. (1965) *Introduction to Probability & Statistics from a Bayesian Viewpoint, Part 2. Inference* (Cambridge University Press).

Mayer, F. (1980) "Specification et choix d'un seuil de signification optimal: Une approche Bayesienne", unpublished doctoral dissertation, Department of Economics, University of Montreal, Montreal, Canada.

Muth, J. (1961) "Rational expectations and the theory of price movements", *Econometrica*, 29, 315–335.

Raiffa, H. and R. Schlaifer (1961) *Applied Statistical Decision Theory* (Graduate School of Business Administration, Harvard University).

Shiller, R.J. (1973) "A distributed lag estimator derived from smoothness priors", *Econometrica*, 41, 775–788.

Winkler, R.L. (1967) "The assessment of prior distributions in Bayesian analysis", *Journal of the American Statistical Association*, 62, 776–800.

Winkler, R.L. (1977) "Prior distributions and model-building in regression analysis", in: A. Aykac and C. Brumat, eds., *New Developments in the Applications of Bayesian Methods* (North-Holland Publishing Co., Amsterdam) pp. 233–242.

Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics* (John Wiley & Sons, Inc., New York).

Zellner, A. (1976) "Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* errors", *Journal of the American Statistical Association*, 71, 400–405.

Zellner, A. (1977) "Maximal data information prior distributions", in: A. Aykac and C. Brumat, eds., *New Developments in the Applications of Bayesian Methods* (North-Holland Publishing Co., Amsterdam), pp. 211–232.

Zellner, A. (1985) "On assessing informative prior distributions for regression coefficients", in: J. M. Bernardo et al., eds., *Bayesian Statistics*, vol. 2 (North-Holland Publishing Co., Amsterdam), pp. 571–581.

Zellner, A. and M.S. Geisel (1970) "Analysis of distributed lag models with applications to consumption function estimation", *Econometrica*, 38, 865–888.

Zellner, A. and J.F. Richard (1973) "Use of prior information in the analysis and estimation of Cobb–Douglas production function models", *International Economic Review*, 14, 107–119.

Zellner, A. and A. Williams (1973) "Bayesian analysis of the federal reserve–MIT–PENN model's Almon lag consumption function", *Journal of Econometrics*, 1, 267–299.

Zellner, A. and W. Vandaele (1975) "Bayes–Stein estimators for *k*-means, regression and simultaneous equation models", in S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* (North-Holland Publishing Co., Amsterdam), pp. 627–653.

Zellner, A. and A. Siow (1979) "Posterior odds ratios for selected regression hypotheses", presented at the International Meeting on Bayesian Statistics, Valencia, 28 May–2 June 1979 and published in J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics* (University Press, Valencia, Spain 1980).