Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 7006

POSTERIOR PROBABILITIES OF ALTERNATIVE LINEAR MODELS

by T. Kloek and F.B. Lempers

February 18, 1970

Preliminary and Confidential

POSTERIOR PROBABILITIES OF ALTERNATIVE LINEAR MODELS

by T. Kloek and F.B. Lempers[1]

Contents

1. INTRODUCTION

One of the most challenging concepts in Bayesian statistics is the probability of a particular model being correct. Discussions of this concept can be found in Jeffreys (1961) and Carnap (1962), and attempts to apply it have been undertaken by Thornber (1966) and by Zellner and Geisel[1a] (1968). A common point in the approach of these authors is that they start from non-informative prior distributions.

This paper treats the same subject starting from natural-conjugate distributions, as introduced by Raiffa and Schlaifer (1961) who, in turn, base their approach on the foundations formulated by Savage (1954). Special attention is given to the case of a set of alternative linear models. The subject is also reconsidered for the case of uniform prior distributions to illustrate some of the difficulties inherent in Jeffreys' approach.

---

1 The authors are indebted to Mr. R. Harkema of the Econometric Institute, who gave some valuable comments.

1a During the stage of proof reading the authors were informed that there is more unpublished work in this area by both Geisel and Zellner. The authors wish to point out that the present results have been derived independently.

Sets of alternative linear models are widely used in applied work (in econometrics for instance), and traditionally[2] the squared multiple correlation coefficient (or determination coefficient) R^2 often plays an important role in choosing from such a set. It can be shown that under certain conditions this device works well for large samples, but it can produce bad results when the number of observations is limited.[3]

In principle, the posterior probability of a linear hypothesis could take over the role of R^2. However, as is argued in Section 7, several problems must be solved before a wide-scale use of this approach can be advocated. This paper should be considered as a preliminary result which we hope will form a basis for further research.

The order of discussion is as follows. In Section 2 the general posterior probability of a hypothesis out of a set of mutually exclusive hypotheses is derived; Sections 3, 4, and 5 apply this to a set of alternative linear hypotheses for natural-conjugate as well as uniform prior distributions. In Section 6, a large sample property is derived and, finally, in Section 7, some comments are made about applicability. The Appendix contains a rather technical proof for a result required in Section 6.

2. POSTERIOR PROBABILITIES OF SUBSETS OF THE STATE SPACE

Let us consider a state space Θ which can be partitioned into a finite[4] number of measurable subspaces $\Theta_1, \ldots, \Theta_k$ as follows:

$$\Theta = \bigcup_{i=1}^{k} \Theta_i \tag{2.1}$$

$$\Theta_i \cap \Theta_j = \emptyset \qquad (i, j = 1, \ldots, k;\ i \neq j) \tag{2.2}$$

We denote the hypothesis that $\theta \in \Theta_i$ by H_i and the probability that H_i is true by p_i so that

$$p_i \equiv P[\theta \in \Theta_i] \tag{2.3}$$

---

[2]See, e.g., Goldberger (1968), Section 9.4.

[3]See Kloek (1970) for a large sample result, and Koerts and Abrahamse (1969), Chapter 8, for small sample results.

[4]The generalization to a countable number is obvious.

We consider the case that all Θ_i are uncountable and that for each i we have a conditional prior density $D'(\theta \mid \Theta_i)$. Let $K'(\theta \mid \Theta_i)$ be a kernel of $D'(\theta \mid \Theta_i)$; that is, let

$$D'(\theta \mid \Theta_i) = C_i' K'(\theta \mid \Theta_i) \tag{2.4}$$

where C_i' is implicitly defined by

$$C_i' \int_{\Theta_i} K'(\theta \mid \Theta_i) d\theta = 1 \tag{2.5}$$

If the condition $\theta \in \Theta_i$ is dropped, the prior[5] is $p_i' D'(\theta \mid \Theta_i)$ $(i = 1, \ldots, k;\ \theta \in \Theta_i)$. Let $\ell_i(z \mid \theta)$ denote the likelihood of the sample $z = (z(1), \ldots, z(n))$ given $\theta(\theta \in \Theta_i)$. This likelihood can be written as

$$\ell_i(z \mid \theta) = \kappa_i(z \mid \theta)\rho_i(z) \tag{2.6}$$

where $\kappa_i(z \mid \theta)$ is a kernel of the likelihood of z given $\theta(\theta \in \Theta_i)$. Then the posterior density of $\theta \in \Theta_i$, given z, reads

$$\begin{aligned} D''(\theta \mid z) &= \frac{p_i' D'(\theta \mid \Theta_i)\ell_i(z \mid \theta)}{\sum_j p_j' \int_{\Theta_j} D'(t \mid \Theta_j)\ell_j(z \mid t)dt} \\ &= \frac{p_i' C_i' \rho_i(z) K'(\theta \mid \Theta_i)\kappa_i(z \mid \theta)}{\sum_j p_j' C_j' \rho_j(z) \int_{\Theta_j} K'(t \mid \Theta_j)\kappa_j(z \mid t)dt} \\ & \qquad (i = 1, \ldots, k;\ \theta \in \Theta_i) \end{aligned} \tag{2.7}$$

and the posterior probability that H_i is true

$$\begin{aligned} p_i'' \equiv P''[\Theta_i \mid z] &= \int_{\Theta_i} D''(\theta \mid z) d\theta = \\ &= \frac{p_i' C_i' \rho_i(z) \int_{\Theta_i} K'(t \mid \Theta_i)\kappa_i(z \mid t)dt}{\sum_j p_j' C_j' \rho_j(z) \int_{\Theta_j} K'(t \mid \Theta_j)\kappa_j(z \mid t)dt} \end{aligned} \tag{2.8}$$

---

[5] Note that this prior is a mixed mass-density function.

Later we shall discuss the ratio of the posterior probabilities of two alternative hypotheses

$$(2.9) \qquad \frac{p_i''}{p_j''} = \frac{p_i'\rho_i(z)C_i'/C_i''(z)}{p_j'\rho_j(z)C_j'/C_j''(z)}$$

where $C_i''(z)$ is defined by

$$(2.10) \qquad C_i''(z) \int_{\Theta_i} K'(t \mid \Theta_i)\kappa_i(z \mid t)dt = 1$$

From (2.9) we conclude that the integration constants C_i' and C_j' of the prior distributions play a crucial role in the analysis.

3. THE CASE OF SEVERAL LINEAR MODELS: THE LIKELIHOOD FUNCTION

In comparing different models to explain one variable, one usually uses the same sample z. The formulation of these models, however, may have been chosen in such a way that the variables to be explained are different; in such a case, they are in general functionally related. Therefore, we introduce the linear model as $y_i = X_i\beta_i + \varepsilon_i$, where $y_i = f_i(z)$ and f_i is a known one-to-one mapping which sends the sample space $Z = \{z\} \subset R^n$ into R^n; well-known examples are $y_i(t) = \log z(t)$ $(t = 1, \ldots, n)$ and $y_i = A_iz$ where A_i is a known non-singular matrix. Furthermore[6] X_i is a fixed matrix of order $n \times r_i$ and rank m_i,[7] and ε_i is a normally distributed random n-vector of disturbances with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i\varepsilon_i^t) = (1/h_i)I$; here t denotes transposition and h_i the precision of the i-th process, that is, the reciprocal of its variance.

Given these assumptions, the likelihood of the sample is

$$(3.1) \qquad \ell_i(z \mid \theta) = \ell(y_i \mid \theta) \ |J_i| = c_i e^{-q_i} h_i^{s_i} \ |J_i| \qquad (\theta \in \Theta_i)$$

---

[6] It may happen that X_i is a submatrix of X_j, which implies $\Theta_i \subset \Theta_j$ contrary to (2.2). This problem can easily be solved by defining a new subspace $\Theta_j^* = \Theta_j - \Theta_i$. Since Θ_j^* contains "almost every" point of Θ_j and since we confine ourselves to continuous distributions on Θ_j, corresponding integrals over Θ_j and Θ_j^* are equal.

[7] The notation in this paper has been taken from Raiffa and Schlaifer (1961), Chapter 13, with four exceptions: (i) since we prefer to use p for probability, m stands for rank; (ii) since we dislike the use of v and ν in the same formula, we use λ rather than ν; (iii) since we prefer to use capitals for matrices, we have replaced n (bold face roman) by N; (iv) no tildes have been used to denote random variables.

with

(3.2a) $$c_i = (2\pi)^{-\frac{1}{2}n}$$

(3.2b) $$q_i = \tfrac{1}{2}h_i\{(\beta_i - b_i)^t N_i(\beta_i - b_i) + \lambda_i v_i\}$$

(3.2c) $$s_i = \tfrac{1}{2}m_i + \tfrac{1}{2}\lambda_i$$

(3.2d) $$|J_i| = |\det[\partial y_i(s)/\partial z(t)]|$$

and

(3.2e) $$\rho_i(z) = c_i\ |J_i| = (2\pi)^{-\frac{1}{2}n}\ |J_i|$$

where

(3.3a) $$N_i = X_i^t X_i$$

(3.3b) $$(X_i^t X_i)b_i = X_i^t y_i$$

(3.3c) $$m_i = \text{rank}\ [N_i]$$

(3.3d) $$\lambda_i = n - m_i$$

(3.3e) $$v_i = (y_i - X_i b_i)^t (y_i - X_i b_i)/\lambda_i$$

see Raiffa and Schlaifer (1961), Chapter 13.

4. PRIOR AND POSTERIOR ANALYSIS WITH NATURAL-CONJUGATE PRIOR DISTRIBUTIONS

In specifying a conditional Natural-conjugate prior density for the parameters of the likelihood function (3.1), we suppose that this prior distribution has been fitted to subjective betting odds.[8] We follow Raiffa and Schlaifer (1961), Chapter 13, and take the Normal-gamma density

(4.1) $$D'(\theta \mid \Theta_i) = D'(\beta_i, h_i \mid \Theta_i) = c_i' e^{-q_i'} h_i^{s_i'}$$

[8] One may remark that betting on the question as to which of two models is the "true" one is difficult since one will never know with certainty whether a model is "true" or not. A possible solution to this problem might be that the player who bets that model i is "true" wins if model i yields better conditional predictions than the other model in a previously specified prediction problem.

with

(4.2a) $$c_i' = (2\pi)^{-\frac{1}{2}m_i'} |N_i'|^{\frac{1}{2}} (\tfrac{1}{2}\lambda_i' v_i')^{\frac{1}{2}\lambda_i'} / (\tfrac{1}{2}\lambda_i' - 1)!$$

(4.2b) $$q_i' = \tfrac{1}{2}h_i\{(\beta_i - b_i')^t N_i'(\beta_i - b_i') + \lambda_i' v_i'\}$$

(4.2c) $$s_i' = \tfrac{1}{2}m_i' + \tfrac{1}{2}\lambda_i' - 1$$

where $m_i' = \text{rank } [N_i']$. In order to guarantee that (4.1) is a proper density, we assume for $i = 1, \ldots, k$ that $\lambda_i' > 0$, $v_i' > 0$, and N_i' is positive-definite symmetric, so that

(4.3) $$m_i' = r_i$$

Combining this prior density with the likelihood function (3.1), we find a Normal-gamma posterior distribution

(4.4) $$D''(\theta \mid z) = c_i'' e^{-q_i''} h_i^{s_i''}$$

with

(4.5a) $$c_i'' = (2\pi)^{-\frac{1}{2}m_i''} |N_i''|^{\frac{1}{2}} (\tfrac{1}{2}\lambda_i'' v_i'')^{\frac{1}{2}\lambda_i''} / (\tfrac{1}{2}\lambda_i'' - 1)!$$

(4.5b) $$q_i'' = \tfrac{1}{2}h_i\{(\beta_i - b_i'')^t N_i''(\beta_i - b_i'') + \lambda_i'' v_i''\}$$

(4.5c) $$s_i'' = \tfrac{1}{2}m_i'' + \tfrac{1}{2}\lambda_i'' - 1$$

where

(4.6a) $$N_i'' = N_i' + N_i$$

(4.6b) $$b_i'' = (N_i'')^{-1}(N_i' b_i' + N_i b_i)$$

(4.6c) $$m_i'' = \text{rank } [N_i'']$$

(4.6d) $$\lambda_i'' = \lambda_i' + m_i' + \lambda_i + m_i - m_i''$$

and[9]

(4.6e) $$v''_i = \{\lambda'_i v'_i + b'^t_i N'_i b'_i + \lambda_i v_i + b^t_i N_i b_i - b''^t_i N''_i b''_i\}/\lambda''_i$$

$$= \{\lambda'_i v'_i + \lambda_i v_i + (b_i - b'_i)^t\{(N'_i)^{-1} + N_i^{-1}\}^{-1}(b_i - b'_i)\}/\lambda''_i$$

and find that the ratio of posterior probabilities (2.9) can now be rewritten as

(4.7) $$\frac{p''_i}{p''_j} = \frac{p'_i|N'_i|^{\frac{1}{2}}|N''_j|^{\frac{1}{2}}(\frac{1}{2}\lambda'_i v'_i)^{\frac{1}{2}\lambda'_i}(\frac{1}{2}\lambda''_j v''_j)^{\frac{1}{2}\lambda''_j}}{p'_j|N'_j|^{\frac{1}{2}}|N''_i|^{\frac{1}{2}}(\frac{1}{2}\lambda'_j v'_j)^{\frac{1}{2}\lambda'_j}(\frac{1}{2}\lambda''_i v''_i)^{\frac{1}{2}\lambda''_i}} \times \frac{|J_i|(\frac{1}{2}\lambda'_j - 1)!(\frac{1}{2}\lambda''_i - 1)!}{|J_j|(\frac{1}{2}\lambda'_i - 1)!(\frac{1}{2}\lambda''_j - 1)!}$$

5. PRIOR AND POSTERIOR ANALYSIS WITH UNIFORM PRIOR DISTRIBUTIONS

In this section we investigate the results of Section 2 for the likelihood function of Section 3 combined with uniform prior distributions for β and log h. We begin by studying the conditional posterior distributions (for $i = 1, \ldots, k$) of β_i and h_i (given that the i-th model is the correct one) and drop for convenience the subscript i. Let R^r stand for an r-dimensional Euclidean space, R^+ denote the set of all positive real numbers, $\{\beta_{\ell\alpha}\}_{\alpha=1}^{\infty}$, $\{\beta_{u\alpha}\}_{\alpha=1}^{\infty}$ denote sequences in R^r, where $\beta_{\ell\alpha} < \beta_{u\alpha}$ ($\alpha = 1, 2, \ldots$), $\{h_{\ell\alpha}\}_{\alpha=1}^{\infty}$ and $\{h_{u\alpha}\}_{\alpha=1}^{\infty}$ denote sequences in R^+, where $h_{\ell\alpha} < h_{u\alpha}$ ($\alpha = 1, 2, \ldots$), and $\{B_\alpha\}_{\alpha=1}^{\infty}$ denote a sequence of subsets ("blocks") of R^{r+1}, where

9 See the Appendix for the derivation of the second equality in (4.6e). Note that in the second line of (4.6e) the assumption is made that N_i has full rank, in other words, that $m_i = r_i$. This assumption is essential in Sections 5 and 6 and in the Appendix, but can be dropped in the rest of this one.

(5.1) $\quad B_\alpha = \{(\beta, h) \mid \beta \in R^r, h \in R^+, \beta_{\ell\alpha} < \beta < \beta_{u\alpha}, h_{\ell\alpha} < h < h_{u\alpha}\}$

Furthermore, let $\{B^*_\alpha\}_{\alpha=1}^{\infty}$ denote a sequence of subsets of R^{r+1}, where

(5.2) $\quad B^*_\alpha = \{(\beta, \log h) \mid \beta \in R^r, h \in R^+, \beta_{\ell\alpha} < \beta < \beta_{u\alpha}, h_{\ell\alpha} < h < h_{u\alpha}\}$

We assume that the components of β and $\log h$ are independently distributed, and specify the following prior densities for $\alpha = 1, 2, \ldots$

(5.3)
$$\begin{aligned} f'_\alpha(\beta, \log h) &= \gamma'_\alpha \qquad \text{if} \quad (\beta, \log h) \in B^*_\alpha \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

with

(5.4)
$$1/\gamma'_\alpha = \{\prod_{j=1}^{r} (\beta_{u\alpha j} - \beta_{\ell\alpha j})\}(\log h_{u\alpha} - \log h_{\ell\alpha})$$

where $\beta_{u\alpha j}$ denotes the j-th component of the r-dimensional vector $\beta_{u\alpha}$, and where $\beta_{\ell\alpha j}$ is similarly defined. Transformation of log h to h yields

(5.5)
$$\begin{aligned} f'_\alpha(\beta, h) &= \gamma'_\alpha h^{-1} \qquad \text{if} \quad (\beta, h) \in B_\alpha \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

Combining the prior densities (5.5) and the likelihood $\ell(y \mid \beta, h)$, we obtain the posterior densities

(5.6)
$$\begin{aligned} f''_\alpha(\beta, h) &= \gamma''_\alpha \frac{|N|^{\frac{1}{2}}(\frac{1}{2}\lambda v)^{\frac{1}{2}\lambda} h^{\frac{1}{2}n-1}}{(2\pi)^{\frac{1}{2}r}(\frac{1}{2}\lambda - 1)!} \exp\,[-\tfrac{1}{2}h\{\lambda v + (\beta - b)^t N(\beta - b)\}] \\ &\qquad\qquad \text{if } (\beta, h) \in B_\alpha \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

where γ''_α is defined by

(5.7)
$$\int_{B_\alpha} f''_\alpha(\beta, h)\, d\beta dh = 1$$

It is clear that γ''_α is approximately equal to unity if $\beta_{u\alpha}$, $-\beta_{\ell\alpha}$, $h_{u\alpha}$, and $1/h_{\ell\alpha}$ are taken large enough. More precisely, the sequence of posterior distributions obtained by considering sequences of values of $\beta_{u\alpha}$, $-\beta_{\ell\alpha}$, $h_{u\alpha}$, and $1/h_{\ell\alpha}$ that diverge to infinity converges in distribution to a Normal-gamma distribution with density

$$f(\beta, h) = f^{(r)}_{N\gamma}(\beta, h \mid b, v, N, \lambda) \tag{5.8}$$

since the limit (5.8) of the sequence of posterior densities so obtained is a proper density; see Scheffé(1947). It follows from (5.5) and (5.6) that we can write for the alternative models

$$C'_i = \gamma'_{\alpha i} \tag{5.9}$$

$$C''_i = \frac{\gamma''_{\alpha i} \, |N_i|^{\frac{1}{2}} (\tfrac{1}{2}\lambda_i v_i)^{\frac{1}{2}\lambda_i}}{(2\pi)^{\frac{1}{2}r_i} (\tfrac{1}{2}\lambda_i - 1)!} \tag{5.10}$$

compare (2.5) and (2.10). These results can be combined with (3.2e) and substituted in (2.9) to obtain

$$\frac{p''_i}{p''_j} = \frac{p'_i \gamma'_{\alpha i} \gamma''_{\alpha j} |N_j|^{\frac{1}{2}} (\tfrac{1}{2}\lambda_i - 1)! (\tfrac{1}{2}\lambda_j v_j)^{\frac{1}{2}\lambda_j} (2\pi)^{\frac{1}{2}r_i} |J_i|}{p'_j \gamma'_{\alpha j} \gamma''_{\alpha i} |N_i|^{\frac{1}{2}} (\tfrac{1}{2}\lambda_j - 1)! (\tfrac{1}{2}\lambda_i v_i)^{\frac{1}{2}\lambda_i} (2\pi)^{\frac{1}{2}r_j} |J_j|} \tag{5.11}$$

If the limiting procedure described above is carried out, we find (for $i = 1, \ldots, k$) that $\gamma''_{\alpha i}$ tends to unity and $\gamma'_{\alpha i}$ tends to zero, which yields an indeterminate result for (5.11) unless additional assumptions are made[10] about the ratio $\gamma'_{\alpha i}/\gamma'_{\alpha j}$. We consider it would be difficult, if not impossible, to defend a general assumption about this ratio, but in certain cases *ad hoc* assumptions might work well.

---

[10] Thornber (1966) adopts a procedure introduced (but rejected) by Jeffreys (1961). His "invariant" prior density is based on the determinant of the so-called information matrix. It has the following shape:

$$\begin{aligned} f_\alpha(\beta_i, h_i) &= \bar{\gamma}'_{\alpha i} |X_i^t X_i|^{\frac{1}{2}} h^{\frac{1}{2}r_i - 1} && \text{if } (\beta_i, h_i) \in B_\alpha \\ &= 0 && \text{otherwise} \end{aligned}$$

He obtains general expressions for the posterior probabilities by assuming (without comment) that $\bar{\gamma}'_{\alpha i} = \bar{\gamma}'_\alpha$ $(i = 1, \ldots, k)$. Note that this prior density generates a marginal posterior distribution on β which is Student with n degrees of freedom independent of the dimension of β.

6. A LARGE SAMPLE PROPERTY

In this section, we leave the field of prior and posterior analysis and consider a property of the conditional sampling distribution of $(p_1'', \ldots, p_k'')$, given the true state θ. For simplicity, we confine ourselves to the probability limit of $(p_1'', \ldots, p_k'')$, as the number of observations tends to infinity. We shall prove that if $\bar{\theta}$ is the true state of nature and belongs to Θ_i, and if

$$|J_j| = |J_i| \qquad (j = 1, \ldots, k)$$

we have

$$\operatorname*{plim}_{n\to\infty} p_i'' = 1 \qquad \operatorname*{plim}_{n\to\infty} p_j'' = 0 \qquad (j \neq i)$$

Without loss of generality, we assume throughout this section that model 1 is the right model and that some unknown pair $(\bar{\beta}_1, \bar{h}_1)$ represents the true state of nature, and we compare this model with some arbitrary model 2. We adopt the assumptions made in the first paragraph of Section 3, but confine ourselves to the case where the variables to be explained are equal for both models, so that $|J_2| = |J_1|$. Further, we need some assumptions about the behavior of the matrices of explanatory variables. Let us write X_{in} for the matrix X_i with n rows, and assume that X_{in} has full column rank (so that $(X_{in}^t X_{in})^{-1}$ exists) and that

$$\lim_{n\to\infty} (X_{in}^t X_{in})^{-1} = 0 \qquad (i = 1, 2) \tag{6.1}$$

In addition, we assume that

$$\lim_{n\to\infty} \frac{|X_{1n}^t X_{1n}|^{\frac{1}{2}}}{|X_{2n}^t X_{2n}|^{\frac{1}{2}}} \cdot \frac{1}{(1+\eta)^n} = 0 \tag{6.2}$$

for some positive number η to be fixed below. Note that it is sufficient for (6.2) that real numbers $A > 0$ and μ exist such that

$$n^{\mu} \frac{|X_{1n}^t X_{1n}|^{\frac{1}{2}}}{|X_{2n}^t X_{2n}|^{\frac{1}{2}}} < A$$

for all values of n from some n_0 onward.

Concerning the parameters of the natural-conjugate prior distribution, we remark that they are independent of n so that the probability limit of (4.7), if it exists, is equal to

$$(6.3) \qquad \operatorname*{plim}_{n\to\infty} \frac{p_2''}{p_1''} = C^* \operatorname*{plim}_{n\to\infty} \frac{|N_1''|^{\frac{1}{2}}(\frac{1}{2}\lambda_1''v_1'')^{\frac{1}{2}\lambda_1''}(\frac{1}{2}\lambda_2''-1)!}{|N_2''|^{\frac{1}{2}}(\frac{1}{2}\lambda_2''v_2'')^{\frac{1}{2}\lambda_2''}(\frac{1}{2}\lambda_1''-1)!}$$

where C^* is implicitly defined as

$$(6.4) \qquad C^* = \frac{p_2'|N_2'|^{\frac{1}{2}}(\frac{1}{2}\lambda_2'v_2')^{\frac{1}{2}\lambda_2'}(\frac{1}{2}\lambda_1'-1)!}{p_1'|N_1'|^{\frac{1}{2}}(\frac{1}{2}\lambda_1'v_1')^{\frac{1}{2}\lambda_1'}(\frac{1}{2}\lambda_2'-1)!}$$

First, let us consider the ratio $(\frac{1}{2}\lambda_2''-1)!/(\frac{1}{2}\lambda_1''-1)!$ If we write $\lambda_2'' = \lambda_1'' + 2\omega$, where ω is some unknown real number, we find by Stirling's formula for sufficiently large values of λ_1'' and λ_2''

$$(6.5) \qquad \frac{(\frac{1}{2}\lambda_2''-1)!}{(\frac{1}{2}\lambda_1''-1)!} = \frac{(\frac{1}{2}\lambda_1''+\omega-1)!}{(\frac{1}{2}\lambda_1''-1)!} \approx \frac{\sqrt{\frac{1}{2}\lambda_1''-1+\omega}\,\{(\frac{1}{2}\lambda_1''-1+\omega)/e\}^{\frac{1}{2}\lambda_1''-1+\omega}}{\sqrt{\frac{1}{2}\lambda_1''-1}\quad\{(\frac{1}{2}\lambda_1''-1)/e\}^{\frac{1}{2}\lambda_1''-1}}$$

$$= \sqrt{1+\frac{\omega}{\frac{1}{2}\lambda_1''-1}}\,\{1+\frac{\omega}{\frac{1}{2}\lambda_1''-1}\}^{\frac{1}{2}\lambda_1''-1}\; e^{-\omega}(\tfrac{1}{2}\lambda_1''-1+\omega)^{\omega}$$

see, e.g., Courant (1937), Vol. 1,Chapter VII. If $n \to \infty$, also $\lambda_1'' \to \infty$ and $\lambda_2'' \to \infty$ (see (4.6d) and (3.3d)), so that (6.3) can be rewritten as

$$(6.6) \qquad \operatorname*{plim}_{n\to\infty} \frac{p_2''}{p_1''} = C^* \operatorname*{plim}_{n\to\infty} \frac{|N_1''|^{\frac{1}{2}}(\lambda_1''v_1''/n)^{\frac{1}{2}\lambda_1''}}{|N_2''|^{\frac{1}{2}}(\lambda_2''v_2''/n)^{\frac{1}{2}\lambda_2''}}$$

since

$$(6.7) \qquad \lim_{n\to\infty} \sqrt{1+\frac{\omega}{\frac{1}{2}\lambda_1''-1}}\,\{1+\frac{\omega}{\frac{1}{2}\lambda_1''-1}\}^{\frac{1}{2}\lambda_1''-1}\; e^{-\omega}(\tfrac{1}{2}\lambda_1''-1+\omega)^{\omega}(\tfrac{1}{2}n)^{-\omega} = 1$$

Next, let us consider

$$\text{(6.8)} \qquad \operatorname*{plim}_{n\to\infty} \frac{\lambda_1'' v_1''}{n} = \operatorname*{plim}_{n\to\infty} \frac{\lambda_1' v_1' + b_1'^{t} N_1' b_1' + \lambda_1 v_1 + b_1^{t} N_1 b_1 - b_1''^{t} N_1'' b_1''}{n}$$

$$= \operatorname*{plim}_{n\to\infty} \frac{\lambda_1 v_1}{n} + \operatorname*{plim}_{n\to\infty} \frac{(b_1 - b_1')^t \{(N_1')^{-1} + N_1^{-1}\}^{-1} (b_1 - b_1')}{n}$$

compare (4.6e). Recalling that $y_n = X_{1n}\bar{\beta}_1 + \varepsilon_n$, where ε_n is normally distributed with zero mean and covariance matrix $(1/\bar{h}_1)I_n$, and $b_1 = (X_{1n}^t X_{1n})^{-1} X_{1n}^t y_n$, it is well-known that

$$\text{(6.9)} \qquad \operatorname*{plim}_{n\to\infty} \frac{\lambda_1 v_1}{n} = \sigma^2$$

where $\sigma^2 = 1/\bar{h}_1$; see, e.g., Mood and Graybill (1963), p. 348. To find the probability limit of the second term of the final expression of (6.8), we start with b_1 and conclude that (6.1) is a necessary and sufficient condition for b_1 to converge in the squared mean to $\bar{\beta}_1$, since $Eb_1 = \bar{\beta}_1$ for each n, and

$$\text{(6.9)} \qquad \lim_{n\to\infty} V(b_1) = \sigma^2 \lim_{n\to\infty} (X_{1n}^t X_{1n})^{-1} = 0$$

Convergence in the squared mean implies convergence in probability, so that

$$\text{(6.10)} \qquad \operatorname*{plim}_{n\to\infty} b_1 = \bar{\beta}_1$$

and

$$\text{(6.11)} \qquad \operatorname*{plim}_{n\to\infty} (b_1 - b_1') = \bar{\beta}_1 - b_1'$$

Next, given assumption (6.1), it is easy to verify that

$$\text{(6.12)} \qquad \lim_{n\to\infty} \{(N_1')^{-1} + N_1^{-1}\}^{-1} = N_1'$$

Combining (6.11) and (6.12), we obtain

$$\text{(6.13)} \qquad \operatorname*{plim}_{n\to\infty} \frac{(b_1 - b_1')^t \{(N_1')^{-1} + N_1^{-1}\}^{-1} (b_1 - b_1')}{n} = 0$$

so that

$$\text{(6.14)} \qquad \underset{n\to\infty}{\text{plim}} \frac{\lambda_1''v_1''}{n} = \sigma^2$$

see (6.8) and (6.9).

Next, we continue with the second model and consider the limiting behavior of

$$\text{(6.15)} \qquad \frac{\lambda_2''v_2''}{n} = \frac{\lambda_2'v_2' + \lambda_2 v_2 + (b_2 - b_2')^t\{(N_2')^{-1} + N_2^{-1}\}^{-1}(b_2 - b_2')}{n}$$

Here $\lambda_2 v_2 = (y_n - X_{2n}b_2)^t(y_n - X_{2n}b_2)$ where $b_2 = (X_{2n}^t X_{2n})^{-1}X_{2n}^t y_n$ while, in fact, y has been generated by $y_n = X_{1n}\bar{\beta}_1 + \varepsilon_n$. We assume that there is no value of β_2 such that $X_{2n}\beta_2 = X_{1n}\bar{\beta}_1$ (otherwise the second specification would be a correct alternative to the first one). Then we can write

$$\text{(6.16)} \qquad \frac{\lambda_2 v_2}{n} = \frac{\bar{\beta}_1^t X_{1n}M_{2n}X_{1n}\bar{\beta}_1 + 2\varepsilon_n^t M_{2n}X_{1n}\bar{\beta}_1 + \varepsilon_n^t M_{2n}\varepsilon_n}{n}$$

where M_{2n} is an idempotent and symmetric matrix defined by

$$\text{(6.17)} \qquad M_{2n} = I - X_{2n}(X_{2n}^t X_{2n})^{-1}X_{2n}^t$$

Our assumption $X_{2n}\beta_2 \neq X_{1n}\bar{\beta}_1$ (all β_2) implies that $M_{2n}X_{1n}\bar{\beta}_1 \neq 0$. Now, if we assume[11] that, as the sample size increases, the sequence $(\bar{\beta}_1^t X_{1n}^t M_{2n}X_{1n}\bar{\beta}_1)/n$ is bounded below by a positive number a_1, it can be shown [see Kloek (1970)] that for each $\delta > 0$

$$\lim_{n\to\infty} P[\frac{\lambda_2 v_2}{n} \geq \sigma^2 + a_1 - \delta] = 1$$

If we choose an arbitrary value for δ such that $0 < \delta < a_1$ and define $a_2 = a_1 - \delta$, we obtain

$$\text{(6.18)} \qquad \lim_{n\to\infty} P[\frac{\lambda_2 v_2}{n} \geq \sigma^2 + a_2] = 1$$

---

[11] Note that $M_{2n}X_{1n}\bar{\beta}_1$ can be interpreted as the vector of least-squares residuals which results when $X_{1n}\bar{\beta}_1$ is "explained" by the columns of X_{2n}. So the assumption says that the mean-square of these residuals will never be smaller than a positive number a.

Concerning the quadratic form on the right hand-side of (6.15), we remark that we do not need an extensive analysis as was the case with model 1, since the matrix $\{(N_2')^{-1} + N_2^{-1}\}^{-1}$ is positive definite (also in the limiting case) and thus the quadratic form is positive for all n. This implies that

$$(6.19) \qquad \lim_{n\to\infty} P[\frac{\lambda_2'' v_2''}{n} \geq \sigma^2 + a_2] = 1$$

see (6.15) and (6.18).

We go on to investigate the ratio of determinants of (6.6) and consider

$$(6.20) \qquad \frac{|N_1''|^{\frac{1}{2}}}{|N_2''|^{\frac{1}{2}}} = \frac{|N_1' + N_1|^{\frac{1}{2}}}{|N_2' + N_2|^{\frac{1}{2}}}$$

$$= \frac{|X_{1n}^t X_{1n}|^{\frac{1}{2}} |(X_{1n}^t X_{1n})^{-1} N_1' + I|^{\frac{1}{2}}}{|X_{2n}^t X_{2n}|^{\frac{1}{2}} |(X_{2n}^t X_{2n})^{-1} N_2' + I|^{\frac{1}{2}}}$$

Note that, if $n \to \infty$, the right-hand determinants in the numerator as well as in the denominator tend to 1 since the corresponding matrices converge to unity matrices; compare (6.1). Collecting (6.14), (6.19), and (6.20), we obtain for (6.6)

$$(6.21) \qquad \operatorname*{plim}_{n\to\infty} \frac{p_2''}{p_1''} \leq C^* \lim_{n\to\infty} \frac{|X_{1n}^t X_{1n}|^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}\lambda_1''}}{|X_{2n}^t X_{2n}|^{\frac{1}{2}} (\sigma^2 + a_2)^{\frac{1}{2}\lambda_2''}}$$

$$= C^* \cdot \sigma^{-2\omega} \lim_{n\to\infty} \frac{|X_{1n}^t X_{1n}|^{\frac{1}{2}}}{|X_{2n}^t X_{2n}|^{\frac{1}{2}}} \cdot \frac{1}{(1 + a_2/\sigma^2)^{\frac{1}{2}\lambda_2''}}$$

$$= 0$$

if (6.2) is assumed to hold true for some $\eta \geq a_2/\sigma^2$. This implies

$$(6.22) \qquad \operatorname*{plim}_{n\to\infty} p_1'' = 1 \quad \text{and} \quad \operatorname*{plim}_{n\to\infty} p_2'' = 0 \ ; \qquad \text{q.e.d.}$$

By investigating the limiting behavior of (5.11) and accepting the assumption that the ratio $\gamma_{\alpha 2}'/\gamma_{\alpha 1}'$ is bounded, we can easily verify the fact that the previous assumptions and results can be used to find that (6.22) also holds if we start from (5.11).

7. REMARKS ABOUT APPLICABILITY

So far we have derived a number of mathematical results. It is too early to discuss fully whether and to what extent these results are applicable. We shall therefore confine ourselves to four remarks.

(1) Of course, the large-sample results in Section 6 imply that for any values of the prior parameters our method will select with near-certainty the correct model - provided there is one in the set of models considered - as long as our sample size is sufficiently large. It should be noted that we have confined ourselves in Section 6 to the comparison of models with the same variable to be explained. Ratios of Jacobians unequal to unity are much more difficult to handle.

(2) If the investigator is able and willing to specify proper natural-conjugate prior densities on all the models he considers, he can (in principle) apply (4.7) without any difficulty. He should be warned, however, that this specification problem is not at all simple. Suppose, for example, that he wants to compare two models which he considers equally probable <u>a priori</u>. He may express this opinion by setting $p_1' = p_2' = \frac{1}{2}$. But then his final result p_1''/p_2'' may be quite sensitive to the way he specifies b_i', N_i', λ_i' and v_i' $(i = 1, 2)$; compare (4.6) and (4.7). So we advise anyone who uses (4.7) to carry out some sensitivity analysis for acceptable variations of the prior parameters. Some of the most obvious possibilities for practical application are discussed in Remark (4) below.

(3) In case the investigator has no information (or very little) on β and h, he may wish to consider the limiting behavior of (5.11) as α tends to infinity. As it has already been mentioned, it will turn out that no unique limit of $\gamma_{\alpha i}'/\gamma_{\alpha j}'$ can be obtained unless a specific assumption is made about "how fast" these quantities tend to zero in comparison with each other. It is impossible to give general rules here, but there are some particular cases in which rather natural solutions are available. For these we refer to the following remark.

(4) We now briefly describe some alternative models for which the comparability problems for both natural-conjugate and uniform prior distributions are relatively small. Let, for example, the models differ only as to alternative assumptions on the disturbances, such as homoskedasticity versus heteroskedasticity, or independent versus Markov autocorrelated disturbances. In such cases, it seems acceptable to specify identical prior distributions on β, and no serious problems arise in specifying the prior parameters pertaining to the precision. Another example for which we hope

to solve the problems of specifying the prior parameters in a consistent way is that of a set of distributed lag models, where the most probable lag parameter must be found. As a third example, we mention the comparison of alternative proxy-variables for explanatory variables which can not be measured. Also in such a case one may often assume that the prior parameters are independent of the way in which the explanatory variables are approximated. For an example of such a problem, we refer to a recent paper by Harkema and Kloek (1969), where this subject is treated by means of the natural-conjugate approach. This example also illustrates that, for purposes of prediction and decision, one is not forced to choose a single model, but may proceed with a mixture of all or some conditional posterior distributions.

REFERENCES

Carnap, R. (1962), Logical Foundations of Probability, The University of Chicago Press, Second Edition.

Courant, R. (1937), Differential and Integral Calculus, Vol. I, Blackie, London, Second Edition.

Goldberger, A.S. (1968), Topics in Regression Analysis, Macmillan, Collier-Macmillan, London.

Harkema, R., and T. Kloek (1969), "An Example of an Informative Prior Distribution in a Decision-Making Context". Report 6913 of the Econometric Institute, Netherlands School of Economics.

Jeffreys, H. (1961), Theory of Probability, Oxford University Press, London, Third Revised Edition.

Kloek, T. (1970), "Note on Consistent Estimation of the Variance of the Disturbances in the Linear Model", Report 7003 of the Econometric Institute, Netherlands School of Economics.

Koerts, J. and A.P.J. Abrahamse (1969), On the Theory and Applications of the General Linear Model, Rotterdam University Press, Rotterdam.

Mood, A.M., and F.A. Graybill (1963), Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., New York, Second Edition.

Savage, L.J. (1954), The Foundations of Statistics, Wiley, New York.

Scheffé, H. (1947), "A Useful Convergence Theorem for Probability Distributions", Annals of Mathematical Statistics, Vol. 18, pp. 434-438.

Thornber, E.H. (1966), "Applications of Decision Theory to Econometrics", Unpublished Doctoral Dissertation, University of Chicago.

Zellner, A., and M.S. Geisel (1968), "Analysis of Distributed Lag Models with Applications to Consumption Function Estimation", Paper presented at the European Meeting on Statistics, Econometrics, and Management Science, Amsterdam 1968, University of Chicago.

A P P E N D I X

In this appendix we derive the second equality of (4.6e). We neglect the subscripts, since the result holds for any of the models considered provided the N matrix has full rank. We make repeated use of the following property, which holds for arbitrary non-singular matrices A and B of the same order:

$$(A.1) \qquad (A + B)^{-1} = B^{-1}(B^{-1} + A^{-1})^{-1}A^{-1} = A^{-1}(A^{-1} + B^{-1})^{-1}B^{-1}$$

Our starting point is $\lambda''v''$ and we find from (4.6e) that

$$(A.2) \qquad \lambda''v'' = \lambda'v' + b'^tN'b' + \lambda v + b^tNb - b''^tN''b''$$

We proceed by rewriting the terms of (A.2) containing b, b', and b", as follows

$$(A.3) \qquad b'^tN'b' + b^tNb - b''^tN''b''$$

$$= b'^tN'b' + b^tNb - (N'b' + Nb)^t(N' + N)^{-1}(N'b' + Nb)$$

$$= b'^tN'b' + b^tNb - (N'b' + Nb)^t(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}(N'b' + Nb)$$

$$= b'^tN'b' + b^tNb - \{b' + (N')^{-1}Nb\}^t\{(N')^{-1} + N^{-1}\}^{-1}(N^{-1}N'b' + b)$$

$$= b'^tN'b' + b^tNb - b'^t\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}N'b' - b^tN(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}b$$

$$- b'^t\{(N')^{-1} + N^{-1}\}^{-1}b - b^tN(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}N'b'$$

$$= T_1 + T_2 - T_3$$

where

$$(A.4) \qquad T_1 = b'^tN'b' - b'^t\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}N'b'$$

$$(A.5) \qquad T_2 = b^tNb - b^tN(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}b$$

$$(A.6) \qquad T_3 = b'^t\{(N')^{-1} + N^{-1}\}^{-1}b - b^tN(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}N'b'$$

We continue to apply (A.1) to obtain

$$\text{(A.7)}\qquad T_3 \equiv b'^t\{(N')^{-1} + N^{-1}\}^{-1}b + b^tN(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}N'b'$$

$$= b'^t\{(N')^{-1} + N^{-1}\}^{-1}b + b^tN(N' + N)^{-1}N'b'$$

$$= b'^t\{(N')^{-1} + N^{-1}\}^{-1}b + b^tNN^{-1}\{(N')^{-1} + N^{-1}\}^{-1}(N')^{-1}N'b'$$

$$= 2b^t\{(N')^{-1} + N^{-1}\}^{-1}b'$$

Next we consider T_1, and find

$$\text{(A.8)}\qquad T_1 \equiv b'^tN'b' - b'^t\{(N')^{-1} + N^{-1}\}^{-1}N^{-1}N'b'$$

$$= b'^tN'b' + b'^t\{(N')^{-1} + N^{-1}\}^{-1}b' - b'^t\{(N')^{-1} + N^{-1}\}^{-1}(N^{-1}N' + I)b'$$

$$= b'^tN'b'+b'^t\{(N')^{-1}+N^{-1}\}^{-1}b'-b'^t\{(N')^{-1}+N^{-1}\}^{-1}\{N^{-1}+(N')^{-1}\}N'b'$$

$$= b'^t\{(N')^{-1} + N^{-1}\}^{-1}b'$$

and in the same way

$$\text{(A.9)}\qquad T_2 \equiv b^tNb - b^tN(N')^{-1}\{(N')^{-1} + N^{-1}\}^{-1}b$$

$$= b^t\{(N')^{-1} + N^{-1}\}^{-1}b$$

Combining (A.7), (A.8), and (A.9), we obtain for (A.3)

$$\text{(A.10)}\qquad b'^tN'b' + b^tNb - b''^tN''b'' =$$

$$= b'^t\{(N')^{-1} + N^{-1}\}^{-1}b' + b^t\{(N')^{-1} + N^{-1}\}^{-1}b - 2b^t\{(N')^{-1} + N^{-1}\}^{-1}b'$$

$$= (b - b')^t\{(N')^{-1} + N^{-1}\}^{-1}(b - b')$$

Now $\lambda''v''$ (compare (A.2)) can be written as

$$\text{(A.11)}\qquad \lambda''v'' = \lambda'v' + \lambda v + (b - b')^t\{(N')^{-1} + N^{-1}\}^{-1}(b - b')$$

q.e.d.