Posterior Odds Ratios for Selected Regression Hypotheses

A. ZELLNER and A. SIOW
University of Chicago

SUMMARY

Bayesian posterior odds ratios for frequently encountered hypotheses about parameters of the normal linear multiple regression model are derived and discussed. For the particular prior distributions utilized, it is found that the posterior odds ratios can be well approximated by functions that are monotonic in usual sampling theory F statistics. Some implications of this finding and the relation of our work to the pioneering work of Jeffreys and others are considered. Tabulations of odds ratios are provided and discussed.

Keywords: BAYESIAN ODDS RATIOS; HYPOTHESIS TESTING; REGRESSION HYPOTHESES; REGRESSION MODEL.

1. INTRODUCTION

In this paper we derive posterior odds ratios for selected sharp hypotheses which are frequently encountered in regression analysis[1]. Our approach involves use of generalized forms of Jeffreys's prior distributions that he regards as appropriate when there is little previous information, that is "...in the early stages of a subject...," Jeffreys (1967, p. 252). Of course if more information is available, more informative prior distributions can of course be employed as has been done by Dickey (1971, 1975, 1977), Leamer (1978), Zellner (1971, p. 307 ff.) and others. Herein, we shall emphasize the situation in which little is known and, as will be seen resulting posterior odds can be expressed in terms of usual t or F statistics and degrees of freedom. Thus the

1. See Jaynes (1976) for valuable analyses of a number of important practical examples illustrating the need for care in formulating relevant hypotheses and using appropriate techniques in order to obtain sensible results.

results to be presented provide a direct small-sample link between Bayesian posterior odds ratios and non-Bayesian test statistics as in the previous work of Jeffreys (1957, 1967, 1978), Lindley (1957) and others. Also, some large sample connections between Bayesian posterior odds ratios and non-Bayesian large sample test statistics are developed which are special cases of the general results of Lindley (1961) and Schwarz (1978).

Several, including Thornber (1966), Geisel (1970), Geisel and Gaver (1974), Leamer (1978), and Lempers (1971) have considered posterior odds ratios for regression hypotheses when little information is available. Our approach differs from those utilized in these works in that we employ prior distributions different from those employed in these works.

Since our approach is an extension of that originally presented by Jeffreys (1967, Ch.V.), we present a brief review of Jeffreys's related results in Section 2. In Section 3 posterior odds ratios for several important regression hypotheses are derived. Section 4 presents some numerical evaluations of the posterior odds ratios derived in Section 3 while a summary of results and some concluding remarks are given in Section 5.

2. REVIEW OF JEFFREYS'S RESULTS

Jeffreys (1967, Ch.V) has derived posterior odds ratios for a number of important testing problems in which little prior information is available and the issue is whether a parameter's value is equal to zero, a sharp null hypothesis. A sharp null hypothesis of "no effect" is frequently encountered and thus it is important to have an analysis of it. Jeffreys refers to such an analysis as "significance testing" and contrasts it with an estimation approach in which no special value of the parameter, for example zero is singled out for special attention. Also, he (1967, p. 251) points out that his estimation prior probability density function (pdf) for representing "knowing little", for example a uniform prior *pdf* is inappropriate for a significance testing situation in which little is known about a parameter's value[2].

To be specific, consider Jeffreys's (1967, p. 268 ff.) analysis of the normal mean problem,

$$y_i = \lambda + u_i \qquad i = 1,2,\ldots,n \tag{2.1}$$

where the y_i's are observations and the u_i's are unobserved errors assumed independently drawn from a normal population with zero mean and standard

2. In regression analysis when we delimit the number of regressors to be finite, we are obviously using sharp null hypotheses about the values of the coefficients of omitted variables.

deviation σ, $0 < \sigma < \infty$ which has an unknown value. The two hypotheses which Jeffreys considers are:

$$H_i : \lambda = 0 \text{ and } H_2 : \lambda \neq 0. \tag{2.2}$$

As regards prior *pdf*'s, under H_1 Jeffreys utilizes

$$p(\sigma|H_1) \propto 1/\sigma \qquad 0 < \sigma < \infty \tag{2.3}$$

Under H_2, Jeffreys (1967, p. 268) remarks that, "From consideration of similarity it [the prior *pdf* for λ under H_2] must depend on σ, since there, is nothing in the problem except σ to give a scale for λ". His prior under H_2 is

$$p(\lambda,\sigma)d\lambda d\sigma \propto f\left(\frac{\lambda}{\sigma}\right)\frac{d\lambda}{\sigma}\frac{d\sigma}{\sigma} \tag{2.4}$$

where $\int_{-\infty}^{\infty} f(\lambda/\sigma)\, d\lambda/\sigma = 1$. Then with prior odds 1:1, the posterior odds ratio, K_{12} is:

$$K_{12} = \frac{\int_0^\infty \sigma^{-n-1} \exp\{-n(y^2 + \hat{\sigma}^2)/2\sigma^2\} d\sigma}{\int_{-\infty}^{\infty} \int_0^\infty F(\lambda/\sigma)\, \sigma^{-n-2} \exp\{-n[(\lambda-y)^2 + \hat{\sigma}^2]/2\sigma^2\} d\sigma d\lambda} \tag{2.5}$$

where $y = \Sigma_{i=1}^n y_i/n$ and $n\hat{\sigma}^2 = \Sigma_{i=1}^n (y_i-y)^2$.

From detailed consideration of (2.5) in the case $n = 1$ in which no decision regarding H_1 and H_2 can be made ($K_{12} = 1$), Jeffreys finds "that the consideration that one observation shall give an indecisive result is satisfied if $f(v)$ [with $v = \lambda/\sigma$] is any even function with integral 1." (p. 269). Further, the condition that $K_{12}=0$ for $n \geq 2$ when $\hat{\sigma}=0$ and $y \neq 0$ requires that the denominator of (2.5) diverge. This will occur if and only if $\int_0^\infty f(v)v^{n-1}dv$ diverges (p.269). As Jeffreys notes, "the simplest function satisfying this condition for $n > 1$ and also satisfying (3) [$\int_{-\infty}^{\infty} f(v)dv = 1$] is $f(v) = 1/\pi(1 + v^2)$." Thus his form for $f(\lambda/\sigma)$ is

$$f\left(\frac{\lambda}{\sigma}\right)\frac{d\lambda}{\sigma} = \frac{1}{\pi}\,\frac{1}{1+\lambda^2/\sigma^2}\,\frac{d\lambda}{\sigma} \qquad -\infty < \lambda < \infty \tag{2.6}$$

a *pdf* in the univariate Cauchy form centered at zero. With respect to this point, Jeffreys (1967, p.251) states, "We must... say that the mere fact that it has been suggested that λ is zero corresponds to some presumption that it is fairly small". After pointing to unsatisfactory features of a normal prior *pdf* for λ,[3] he writes, "The chief advantage of the form [(2.6)] what we have chosen is that in any significance test it leads to the conclusion that if the null hypothesis [$\lambda = 0$] has a small posterior probability, the posterior probability of the parameters is nearly the same as in the estimation problem. Some difference remains but it is only a trace". (p. 273).

When (2.6) is substituted in (2.5) and the integrations are performed, approximately in terms of the denominator, Jeffreys obtains (1967, p. 272):

$$K_{12} \doteq (\pi\nu/2)^{1/2} \,/\, (1 + t^2/\nu)^{(\nu-1)/2} \tag{2.7}$$

where $\nu = n\text{-}1$ and $t = \sqrt{n}\, y/s$, with $s^2 = \Sigma_{i=1}^{n} (y_i\text{-}y)^2/\nu$ and the error of the approximation "is of the order of $1/n$ of the whole expression". Also Jeffreys (p.274) provides an exact expression for K_{12}. Shown below are values of K_{12} for selected values of ν and t^2 taken from Jeffreys's table (p.439):

From Table 2.1, it is seen that when $\nu = 20$, $K_{12} = 1$ when $t^2 = 4.0$ while for $\nu = 5{,}000$, $K_{12} = 1$ when $t^2 = 9$. It is thus seen that as ν increases in value, a larger value of t^2 is required for indifference ($K_{12} = 1$) between H_1 and H_2. This corresponds to a sampling theorist's usual lowering of the significance level as ν grows in value and also bears a direct relationship to Lindley's Paradox (1957). Also note that in contrast to DeGroot's (1973) result, the tail area or "*p*-value" associated with the *t*-value is *not* equal to the posterior probability on the null hypothesis[4]. For example, with $\nu = 20$ and $t = 2.0$, the "*p*-value" is approximately .025 and yet $K_{12} = 1$ or the posterior probability on $H_1 : \lambda = 0$ is ½. Finally, as Jeffreys (1967, p. 272) remarks, the variation of K_{12} with t is much more important than its variation with ν. For moderately large ν, $K_{12} \doteq (\pi\nu/2)^{1/2}\exp(-t^2/2)$, from which the dependence of K_{12} on ν and t is clearly seen.

In a brief treatment of regression, Jeffreys (1967, pp. 324-326) remarks that "...The whole of the tests related to the normal law of error can be adapted immediately to tests concerning the introduction of a new function to

3. Jeffreys (1967, p.273) points out that if the prior *pdf* for $v = \lambda/\sigma$ were $p(v) \propto \exp\{-cv^2\}$, where c is some given positive constant, the posterior odds ratio for $\lambda = 0$ and $\lambda \neq 0$ "... would never be less than some positive function of n [the sample size] however closely the observations agreed among themselves". Also, on this same page he points out a second defect of this normal form for the prior *pdf*.

4. It appears that DeGroot (1973) obtains his result that the tail area associated with a sampling theory test statistics's value is equal to the posterior probability on the null hypothesis by use of a very special prior *pdf* on his parameter θ. His prior probabilities on θ's possible values are fixed even though a given departure of θ from its null value of zero implies differing departures of the underlying location parameter's value from zero as n, the sample size changes.

TABLE 2.1

Values of t^2 Associated with Corresponding Values of K_{12} and $\nu = n$-1 from (2.7)

ν	K_{12}: 1	$10^{-1/2}$	10^{-1}	$10^{-3/2}$	10^{-2}
9*	3.5	7.7	13.3	...	...
15	3.8	7.1	11.1	15.9	21.5
20	4.0	7.0	10.6	14.5	18.9
50	4.6	7.4	10.0	12.8	16.0
100	5.2	7.7	10.3	12.8	15.5
200	5.7	8.2	10.7	13.1	15.6
500	6.8	9.1	11.4	13.8	16.2
1,000	7.4	9.7	12.0	14.3	16.6
2,000	8.1	10.4	12.7	15.0	17.3
5,000	9.0	11.3	13.6	15.9	18.2
10,000	9.7	12.0	14.3	16.6	18.9
50,000	11.3	13.6	15.9	18.2	20.5
100,000	12.0	14.3	16.6	18.9	21.2

represent a series of measures". (p. 325). He considers the important special case for which the hypothesis is that an added term's coefficient is equal to zero and points out that (2.7) is the approximate posterior odds ratio for this problem where t is the usual t-statistic relating to the added term's coefficient and ν is the degrees of freedom associated with the t-statistic. Below it will be seen that Jeffreys's result is included in our general results as a special case.

3. POSTERIOR ODDS RATIOS FOR SELECTED REGRESSION HYPOTHESES

Let our regression model for the nx1 observation vector $\mathbf{y}$ be:

$$\mathbf{y} = \alpha\iota + X\beta + \mathbf{u} \tag{3.1}$$

* For $\nu = 9$, Jeffreys has used his exact result for K_{12} to compute the following t^2 values: 3.8 for $K_{12} = 1$, 7.7 for $K_{12} = 10^{-1/2}$, and 13.1 for $K_{12} = 10^{-1}$. It is seen that the exact results are in good agreement with the approximate results even though $\nu = 9$ is small. Jeffreys (1967, p. 439) tabulates exact values for $\nu = 1,2,3,\ldots,9$.

where ι is an nx1 vector with all elements equal to one, α and β are a scalar parameter and a kx1 vector of parameters with unknown values, $(\iota{:}X)$ is an nx$(k+1)$ given matrix of rank $k+1$ and **u** is an $n+1$ vector of error terms. It is assumed that the variables in X are measured in terms of deviations from their respective sample means and thus $\iota' X = \mathbf{0}'$. Further, the elements of **u** are assumed independently drawn from a normal population with zero mean and finite variance σ^2 with unknown value.

We initially consider the following two hypotheses:

$$H_1 : \beta = 0, -\infty < \alpha < \infty \text{ and } 0 < \sigma < \infty \tag{3.2}$$

$$H_2 : \beta \neq \mathbf{0}, -\infty < \alpha < \infty \text{ and } 0 < \sigma < \infty. \tag{3.3}$$

The likelihood functions under these two hypotheses are given by:

$$\begin{aligned} p(\mathbf{y}\,|\,\alpha,\sigma,H_1) &\propto \sigma^{-n}\exp\{-(\mathbf{y}-\alpha\iota)'(\mathbf{y}-\alpha\iota)/2\sigma^2\} \\ &\propto \sigma^{-n}\exp\{-[\nu_1 s_1^2 + n(\alpha y)^2]/2\sigma^2\} \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} p(\mathbf{y}\,|\,\alpha,\beta,\sigma,H_2) &\propto \sigma^{-n}\exp\{-(\mathbf{y}-\alpha\iota-X\beta)'(\mathbf{y}-\alpha\iota-X\beta)/2\sigma^2\} \\ &\propto \sigma^{-n}\exp\{-[\nu_2 s_2^2 + n(\alpha-\bar{y})^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]/2\sigma^2\} \end{aligned} \tag{3.5}$$

where the proportionality constant is $(2\pi)^{-n/2}$ in each case,

$$y = \Sigma_{i=1}^{n} y_i/n, \;\; \nu_1 s_1^2 = \Sigma_{i=1}^{n} (y_i y)^2, \;\; \nu_1 = n\text{-}1$$
$$\hat{\beta} = (X'X)^{-1}X'\mathbf{y}, \;\; \nu_2 s_2^2 = (\mathbf{y}y\iota - X\hat{\beta})'(\mathbf{y}y\iota - X\hat{\beta}), \text{ and } \nu_2 = n\text{-}k\text{-}1.$$

The following prior assumptions will be utilized in obtaining a posterior odds ratio. First we place equal prior probabilities of ½ on both hypotheses and thus the prior odds ratio is 1:1. Second, under H_1 we employ a diffuse prior distribution for α and σ, that is,

$$p(\alpha,\sigma\,|\,H_1) \propto 1/\sigma \qquad -\infty < \alpha < \infty \text{ and } 0 < \sigma < \infty. \tag{3.6}$$

Under H_2 we utilize the following prior *pdf*

$$p(\alpha,\beta,\sigma\,|\,H_2) \propto f(\beta\,|\,\sigma)/\sigma \qquad -\infty < \alpha < \infty \text{ and } 0 < \sigma < \infty \tag{3.7a}$$

with

$$f(\beta|\sigma) = c|X'X/n\sigma^2|^{1/2}/(1 + \beta'X'X\beta/n\sigma^2)^{(k+1)/2} \quad -\infty < \beta_i < \infty \quad (3.7b)$$
$$i = 1,2,...k$$

where $c = \Gamma[(k+1)/2]/\pi^{(k+1)/2}$.

In (3.6) and (3.7a) the factors of proportionality are assumed the same. Further, in (3.7b) it has been assumed that the prior *pdf* for β given σ is in the form of a k-dimensional multivariate Cauchy probability density function with zero location vector and matrix $X'X/n$, a matrix suggested by the form of the information matrix.

The posterior odds ratio, K_{12} for H_1 and H_2 with the prior odds ratio 1:1, is:

$$K_{12} = \frac{\int p(\mathbf{y}|\alpha,\sigma,H_1)p(\alpha,\sigma|H_1)d\alpha d\sigma}{\int p(\mathbf{y}|\alpha,\beta,\sigma,H_2)p(\alpha,\beta,\sigma|H_2)d\alpha d\beta d\sigma} \quad (3.8)$$

Explicitly, the integration in the numerator of (3.8) is performed as follows. The integral to be evaluated is:

$$I_N = \int_0^\infty \int_{-\infty}^\infty \sigma^{-(n+1)}\exp\{-[\nu_1 s_1^2 + n(\alpha-\bar{y})^2]/2\sigma^2\}d\alpha d\sigma$$

Using properties of the univariate normal *pdf*, integrate with respect to α to obtain:

$$I_N = (2\pi/n)^{1/2}\int_0^\infty \sigma^{-(\nu_1+1)}\exp\{-\nu_1 s_1^2/2\sigma^2\}d\sigma \quad (3.9a)$$
$$= (2\pi/n)^{1/2}\Gamma(\nu_1/2)(2/\nu_1 s_1^2)^{\nu_1/2}/2$$

where the integration over σ was performed by utilizing well-known properties of the inverted gamma *pdf* —see, e.g. Zellner (1971, p. 371)—.

The integral in the denominator of (3.8), denoted by I_D will be evaluated as follows:

$$I_D = \int f(\beta|\sigma,H_2)\sigma^{-(n+1)}\exp\{-[\nu_2 s_2^2 + n(\alpha-\bar{y})^2$$
$$+ (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]/2\sigma^2\}d\alpha d\beta d\sigma$$

with $f(\beta|\sigma)$ given in (3.7b). First integrate over α using properties of the univariate normal *pdf* to obtain:

$$I_D = (2\pi/n)^{1/2}\int f(\beta|\sigma,H_2)\sigma^{-n}\exp\{-[\nu_2 s_2^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]/2\sigma^2\}d\beta d\sigma.$$

On inserting $f(\beta|\sigma,H_2)$ from (3.7a) and performing the integration with respect to the elements of β approximately,[5]

$$I_D \doteq (2\pi/n)^{1/2}c(2\pi/\nu_2)^{k/2}\int_0^\infty \frac{1}{\sigma^n}\frac{1}{\{1+\hat{\beta}'X'X\hat{\beta}/n\sigma^2\}^{(k+1)/2}}\exp\{-\nu_2 s_2^2/2\sigma^2\}d\sigma$$

Then,

$$I_D \doteq (2\pi/n)^{1/2}c(2\pi/\nu_2)^{k/2}\Gamma(\nu_1/2)1/2(2/\nu_2 s_2^2)^{\nu 1/2}/(1+\hat{\beta}'X'X\hat{\beta}/\nu_2 s_2^2)^{(k+1)/2} \quad (3.9b)$$

where the integration over σ has beeen performed approximately.

Using the results in (3.9a) and (3.9b), the approximate posterior odds ratio for H_1 vs. H_2 is given by:

$$K_{12} \doteq (1/c)(\nu_2/2\pi)^{k/2}(\nu_2 s_2^2/\nu_1 s_1^2)^{\nu 1/2}(1+\hat{\beta}'X'X\hat{\beta}/\nu_2 s_2^2)^{(k+1)/2}$$

$$= a(\nu_2/2)^{k/2}(\nu_2 s_2^2/\nu_1 s_1^2)^{(\nu 2-1)/2} \quad (3.10)$$

with $a = \pi^{1/2}/\Gamma[(k+1)/2]$, since $\nu_2 s_2^2 + \hat{\beta}'X'X\hat{\beta} = \nu_1 s_1^2$. Alternatively, K_{12} in (3.10) can be expressed as:

$$K_{12} \doteq a(\nu_2/2)^{k/2}/[1+(k/\nu_2)F_{k,\nu 2}]^{(\nu 2-1)/2} \quad (3.11)$$

or

$$K_{12} \doteq a(\nu_2/2)^{k/2}(1-R^2)^{(\nu 2-1)/2} \quad (3.12)$$

where $F_{k,\nu_2} = \hat{\beta}'X'X\hat{\beta}/ks_2^2$ and $R^2 = \hat{\beta}'X'X\hat{\beta}/(\nu_2 s_2^2+\hat{\beta}'X'X\hat{\beta})$, the usual "F-statistics" and the squared sample multiple correlation coefficient, respectively. Further, a large sample approximation to $-2\ell n K_{12}$ is given by:

$$-2\ell n K_{12} \doteq \chi_k^2 - k\ell n\nu_2 \quad (3.13)$$

5. This approximate integration can be viewed as finding the mean of $f(\beta|\sigma, H_2)$ a bounded function of β. Cramér (1946, p. 353 ff.) indicates that the error of the approximation is $0(n^{-1})$ in line with Jeffreys's remark cited in Section II. Thus if the posterior odds ratio $K_{12}=I_N/I_D$ and if the integral I_N is evaluated exactly and $I_D = I_D^A + 0(n^{-1})$, where I_D^A is the approximate value of I_D, $K_{12} = I_N/[I_D^A + 0(n^{-1})]$ or $I_N/I_D^A = K_{12}[1+0(n^{-1})]$ and thus the error in using I_N/I_D is $K_{12}.0(n^{-1})$, as pointed out above by Jeffreys.

where $\chi_k^2 = \hat{\beta}'X'\hat{\beta}/s_2^2$.[6]

We now consider a hypothesis relating to a subvector of β in (3.1). Rewrite (3.1) as

$$\mathbf{y} = \alpha\iota + X_1\beta_1 + X_2\beta_2 + \mathbf{u} \tag{3.14}$$

where $X = (X_1{:}X_2)$ with X_1 and X_2 k_1x1 and k_2x1 vectors, respectively and $k_1+k_2=k$. All other assumptions made in connection with (3.1) apply to (3.14). For convenience, we shall reparametrize (3.14) as follows:

$$\mathbf{y} = \alpha\iota + X_1\eta + V\beta_2 + \mathbf{u} \tag{3.15}$$

where $V = [I\text{-}X_1(X_1'X_1)^{-1}X_1']X_2$ and $\eta = \beta_1+(X_1'X_1)^{-1}X_1'X_2\beta_2$. Note that $X_1'V = 0$

A posterior odds ratio relating to the following two hypotheses will be derived:

$$H_A : \beta_2 = \mathbf{0} \tag{3.16}$$

and

$$H_B : \beta_2 \neq \mathbf{0} \tag{3.17}$$

with α, the elements of η and σ unrestricted under both hypotheses.

The likelihood functions under these two hypotheses are given by:

$$p(\mathbf{y}\,|\,\alpha,\sigma,\eta,\mathrm{H}_A) \propto \sigma^{-n}\exp\{\nu_A s_A^2 + n(\alpha\text{-}\bar{y})^2 + (\eta\text{-}\hat{\eta})'X_1'X_1(\eta\text{-}\hat{\eta})]/2\sigma^2\} \tag{3.18}$$

and

$$p(\mathbf{y}\,|\,\alpha,\sigma,\eta,\beta_2,H_B) \propto \sigma^{-n}\exp\{\text{-}[\nu_B s_B^2 + n(\alpha\text{-}\bar{y})^2 + (\eta\text{-}\hat{\eta})'X_1'X_1(\eta\text{-}\hat{\eta}) + (\beta_2\text{-}\hat{\beta}_2)'V'V(\beta_2\text{-}\hat{\beta}_2)]/2\sigma^2\} \tag{3.19}$$

where

6. To obtain (3.13), write (3.11) as $K_{12} \doteq a(\nu_2/2)^{k/2}\exp\{\nu^2\text{-}1)/2\,\ell n\,[1+(k/\nu^2)F_{k,\nu_2}]\}$ and expand the logarithmic factor in the exponential as $\ell n(1+x) \doteq x$. The result is $K_{12} \doteq a(\nu_2/2)^{k/2}\exp\{\text{-}kF_{k,\nu_2}/2\}$. Then $\text{-}2\ell nK_{12} \doteq \chi_k^2 - k\ell n\nu_2$, where $\chi_k^2 = kF_{k,\nu_2}$ and terms not depending on ν_2 have been dropped in this large-ν_2 approximation. Further, under $\beta=0$ the approximate sampling *pdf* for $\text{-}2\ell nK_{12}$ can be obtained from that of χ_k^2. Also, again under $\beta = \mathbf{0}$ the approximate cumulative sampling *pdf* for K_{12} in (3.11) can be obtained from that of F_{k,ν_2}. That is, since K_{12} is a one-to-one monotonic function of F_{k,ν_2} for fixed k and ν_2, $\Pr(F_{k,\nu_2} > x) = \Pr(K_{12} < x')$, where x' is the value of K_{12} associated with $F_{k,\nu_2} = x$.

$$\bar{y} = \Sigma_{i=1}^{n} y_i/n, \; \hat{\eta} = (X_1'X_1)^{-1}X_1'\mathrm{y}, \qquad \nu_A s_A^2 = (\mathrm{y}-\bar{y}\iota - X_1\hat{\eta})'(\mathrm{y}-\bar{y}\iota - X_1\hat{\eta}),$$

$$\nu_A = n-k_1-1, \; \hat{\beta}_2 = (V'V)^{-1}V'\mathrm{y}, \quad \nu_B s_B^2 = (\mathrm{y}-\bar{y}\iota - X_1\hat{\eta} - V\hat{\beta}_2)'(\mathrm{y}-\bar{y}\iota - X_1\hat{\eta} - V\hat{\beta}_2)$$

and $\nu_B = n-k_1-k_2-1$.

Under H_A, we employ the following diffuse prior *pdf* for the parameters:

$$p_A(\alpha,\sigma,\eta \,|\, H_A) \propto |X_1'X_1|^{1/2}/\sigma \quad -\infty < \alpha,\eta_i < \infty \quad i=1,2,\ldots,k_1 \tag{3.20}$$

$$0 < \sigma < \infty$$

while under H_B the prior *pdf* is:

$$p_B(\alpha,\sigma,\eta,\beta_2 \,|\, H_B) \propto |X_1'X_1|^{1/2} h(\beta_2 \,|\, \sigma)/\sigma \tag{3.21a}$$

$$-\infty < \alpha, \; \eta_i < \infty, \qquad i=1,2,\ldots,k_1$$
$$0 < \sigma < \infty$$

with

$$h(\beta_2 \,|\, \sigma) = c_B |V'V/n\sigma^2|^{1/2}/(1+\beta_2'V'V\beta_2/n\sigma^2)^{(k_2+1)/2} \tag{3.21b}$$

$$-\infty < \beta_{2i} < \infty \qquad i=1,2,\ldots,k_2.$$

In (3.20) and (3.21a) the factor of proportionality is taken to be the same. In (3.21b), the prior *pdf* for β_2 given σ is the form of a k_2-dimensional multivariate Cauchy *pdf* with zero location vector and matrix $V'V/n$, a matrix suggested by the form of the information matrix.

The posterior odds ratio, K_{AB} for H_A and H_B with the prior odds ratio 1:1 is:

$$K_{AB} = \frac{\int p(\mathrm{y} \,|\, \alpha,\sigma,\eta,H_A) p_A(\alpha,\sigma,\eta \,|\, H_A) d\alpha d\sigma d\eta}{\int p(\mathrm{y} \,|\, \alpha,\sigma,\eta,\beta_2,H_B) p_B(\alpha,\sigma,\eta,\beta_2 \,|\, H_B) d\alpha d\sigma d\eta d\beta_2} \tag{3.22}$$

On applying integration techniques similar to those employed above (see Appendix), the following approximate expression for K_{AB} is obtained:

$$\begin{aligned} K_{AB} &\doteq b(\nu_B/2)^{k_2/2} (\nu_B s_B^2/\nu_A s_A^2)^{(\nu_B-1)/2} \\ &= b(\nu_B/2)^{k_2/2} [(1-R_B^2)/(1-R_A^2)]^{(\nu_B-1)/2} \\ &= b(\nu_B/2)^{k_2/2}/[1+(k_2/\nu_B)F_{k_2\nu_B}]^{(\nu_B-1)/2} \end{aligned} \tag{3.23}$$

where $b = \pi^{1/2}/\Gamma[(k_2+1)/2]$, R_A^2 and R_B^2 are the squared sample multiple correlation coefficientes under H_A and H_B and $F_{k_2\nu_B} = \hat{\beta}_2' V' V\hat{\beta}_2/k_2 s_B^2$, the usual "$F$-statistic". Also, if ν_B is large, the following approximate result is available:

$$-2\ell n K_{AB} \doteq \chi^2_{k_2} - k_2 \ell n \nu_B = \nu_B(R_B^2 - R_A^2)/(1-R_B^2) - k_2 \ell n \nu_B \tag{3.24}$$

with $\chi^2_{k_2} = \hat{\beta}_2' V' V\hat{\beta}_2/s_B^2$.

We now consider the following four hypotheses relating to β_1 and β_2 in (3.14), each assumed to have the same prior probability:

$$H_1 : \beta_1 = \mathbf{0} \text{ and } \beta_2 = \mathbf{0}, \tag{3.25a}$$

$$H_2 : \beta_1 \neq \mathbf{0} \text{ and } \beta_2 \neq \mathbf{0}, \tag{3.25b}$$

$$H_3 : \beta_1 \neq \mathbf{0} \text{ and } \beta_2 = \mathbf{0}, \tag{3.25c}$$

and

$$H_4 : \beta_1 = \mathbf{0} \text{ and } \beta_2 \neq \mathbf{0}, \tag{3.25d}$$

The posterior odds ratio for H_1 and H_2, K_{12}, given in (3.11) is:

$$K_{12} \doteq a(\nu_2/2)^{k/2}/[1 + (k/\nu_2)F_{k\nu_2}]^{(\nu_2-1)/2}, \tag{3.26}$$

where $a = \pi^{1/2}/\Gamma[(k+1)/2]$ and $\nu_2 = n$-k-1. This odds ratio has been derived employing the prior assumptions in (3.6) and (3.7), the latter involving a multivariate Cauchy prior *pdf* for β_1 and β_2 given σ. The posterior odds ratio for H_3 and H_2, K_{32} is identical to K_{AB} in (3.23), namely

$$K_{32} \doteq b(\nu_2/2)^{k_2/2}/[1 + (k^2/\nu_2)F_{k_2\nu_2}]^{(\nu_2-1)/2} \tag{3.27}$$

where $b = \pi^{1/2}/\Gamma[(k_2+1)/2]$ and $\nu_2 = \nu_B = n$-k-1. K_{32} also can be obtained by using the conditional prior *pdf* for β_1 given $\beta_2 = \mathbf{0}$ and σ associated with the multivariate Cauchy *pdf* in (3.7b) under H_2 along with uniform independent priors for α and log σ. Similarly, the posterior odds ratio for H_4 and H_2,K_{42} can be obtained and is:

$$K_{42} \doteq q(\nu_2/2)^{k_1/2}/[1 + (k_1/\nu_2)F_{k_1\nu_2}]^{(\nu_2-1)/2} \tag{3.28}$$

where $q = \pi^{1/2}/\Gamma[(k_1+1)/2]$. Last, from (3.27) and (3.28), the posterior odds ratio for H_3 and H_4,K_{34} is:

$$K_{34} = K_{32}/K_{42} \doteq g(\nu_2/2)^{(k_2-k_1)/2} \left(\frac{1+(k_1/\nu_2)F_{k_1\nu_2}}{1+(k_2/\nu_2)F_{k_2\nu_2}} \right)^{(\nu_2-1)/2} \tag{3.29}$$

where $g = \Gamma[(k_1+1)/2]/\Gamma[(k_2+1)/2]$.

The posterior odds ratios in (3.26)-(3.29) can be helpful in screening sets of variables, X_1 and X_2 for inclusion a in regression in situations in which there is little prior information and the initial presumption is that neither set of variables probably belongs in the regression. A special case of the above analysis is one in which X_1 and X_2 are vectors and thus β_1 and β_2 are scalars. In this case, we are screening individual variables por possible inclussion in the regression. Further, elaboration of the hypotheses in (3.25) to relate individual coefficients is possible and would lead to posterior odds ratios useful in determining which individual variables to include in a regression.

To gain greater familiarity with the odds ratios derived above, we now turn to consider some numerical evaluations of them.

4. NUMERICAL EVALUATION OF SELECTED ODDS RATIOS

In this Section, we provide some numerical evaluations of the odds ratios derived in Section III. First, note that when $k = 1$, the posterior odds ratio K_{12} in (3.11) for the hypotheses $\beta = \mathbf{0}$ and $\beta \neq \mathbf{0}$ reduces to $K_{12} \doteq (\pi\nu_2/2^{1/2}/(1 + t^2/\nu^2)^{(\nu_2-1)/2}$, with $\nu^2 = n - 2$ which is exactly in the form of Jeffreys's odds ratio in (2.7). Thus the numerical results in Table 2.1 apply directly to the case of simple regression. From Table 2.1, it is seen that for $\nu_2 = 20$, $K_{12} \doteq 1$ when $t^2 = 4.0$ where $t^2 = \hat{\beta}^2\Sigma(x_i - \bar{x})^2/s^2$ is the square of the usual t-statistic. Since $r^2 = t^2/(\nu^2 + t^2)$, a value of $r^2 = 1/6$ corresponds to $t^2 = 4.0$ and $K_{12} \doteq 1$ for $\nu_2 = 20$. For $\nu_2 = 5{,}000$ and $t^2 = 9.0$ (or $r^2 = .0018$), $K_{12} \doteq 1$. Thus indifference, ($K_{12} \doteq 1$) is achieved for a larger value of t^2 (or a lower value of r^2) with $\nu_2 = 5{,}000$ as compared with $\nu_2 = 20$. For $\nu_2 = 20$, $K_{12} \doteq 1/100$, that is the odds are 100:1 against $\beta = 0$ when $t^2 = 18.9$ or $r^2 = .486$. For $\nu_2 = 5{,}000$, $K_{12} \doteq 1/100$ when $t^2 = 18.2$ or $r^2 = .00377$. Thus with $\nu_2 = 5{,}000$, a value of $t^2 = 18.2$ (or equivalently, $r^2 = .00377$) strongly favors the hypotheses $\beta \neq 0$. Since values of ν_2 in the vicinity of several thousand are frequently encountered in analyses of cross-section or survey data, these results are relevant for applied work. In particular, they point (a) the need for absolutely larger t-values for indifference ($K_{12} = 1$) as ν_2 increases and (b) recognition that for large values of ν_2, small values of r^2 can be consistent with strong evidence *against* $\beta = 0$. These results, it must be emphasized, apply in situations in which we have little prior information about β's value under the hypotheses $\beta \neq 0$. If more information is available, suitable prior *pdf*'s reflecting it would have to be introduced, as pointed out by Jeffreys (1967, p. 252).

TABLE 4.1
Values of R^2 and F_{k,ν_2} Associated with Particular Values of K_{12} and k in (3.12) for $\nu_2 = 20$ and $\nu_2 = 100$*

A. $\nu_2 = 20$

k	Value of:	K_{12}: 1	$10^{-1/2}$	10^{-1}	$10^{-3/2}$	10^{-2}	.01 and .05 Critical Values of F and Associated R^2's: .01	.05
1	R^2	.16	.26	.35	.42	.49	.29	.18
	$F_{1,20}$	4.0	7.0	10.6	14.5	18.9	8.10	4.35
2	R^2	.27	.35	.43	.49	.55	.37	.26
	$F_{2,20}$	3.7	5.5	7.5	9.7	12.3	5.85	3.49
3	R^2	.35	.42	.48	.54	.60	.43	.32
	$F_{3,20}$	3.5	4.8	6.3	8.0	9.9	4.94	3.10
4	R^2	.40	.47	.53	.58	.63	.47	.36
	$F_{4,20}$	3.4	4.4	5.7	7.0	8.6	4.43	2.87
5	R^2	.45	.51	.57	.61	.66	.51	.40
	$F_{5,20}$	3.2	4.2	5.2	6.4	7.7	4.10	2.71
6	R^2	.48	.54	.59	.64	.68	.54	.44
	$F_{6,20}$	3.1	3.9	4.9	5.9	7.1	3.87	2.60

B. $\nu_2 = 100$

k	Value of:	K_{12}: 1	$10^{-1/2}$	10^{-1}	$10^{-3/2}$	10^{-2}	.01	.05
1	R^2	.050	.072	.093	.11	.13	.065	.038
	$F_{1,100}$	5.2	7.7	10.3	12.8	15.5	6.90	3.94
2	R^2	.089	.11	.13	.15	.17	.088	.058
	$F_{2,100}$	4.9	6.2	7.5	8.8	10.3	4.82	3.09
3	R^2	.12	.14	.16	.18	.20	.11	.075
	$F_{3,100}$	4.6	5.5	6.4	7.4	8.3	3.98	2.70
4	R^2	.15	.17	.19	.21	.23	.12	.090
	$F_{4,100}$	4.4	5.1	5.9	6.6	7.3	3.51	2.46
5	R^2	.18	.20	.21	.23	.25	.14	.10
	$F_{5,100}$	4.3	4.9	5.5	6.1	6.7	3.20	2.30
6	R^2	.20	.22	.24	.25	.27	.15	.12
	$F_{6,100}$	4.2	4.7	5.2	5.7	6.2	2.99	2.19

*Note that $F_{k,\nu_2} = (\nu_2/k)R^2/(1-R^2)$, with $\nu_2 = n$-k-1.

In Table 4.1, we have evaluated the posterior odds ratio K_{12} for $H_1 : \beta = 0$ vs. $H_2 : \beta \neq 0$ for $\nu_2 = 20$ and $\nu_2 = 100$, where $\nu_2 = n\text{-}k\text{-}1$ for selected values of k, the number of elements in β and selected values of R^2, the sample squared multiple correlation coefficient. Also shown in the table are values of associated F-statistics, $F_{k,\nu_2} = (\nu_2/k)R^2/(1\text{-}R^2)$, and .01 and .05 critical values of the F statistic as well as the R^2 values associated with these critical values. From the first line of Table 4.1, we see that for $k = 1$ and $\nu_2 = n\text{-}k\text{-}1 = 20$, $K_{12} \doteq 1$ when $R^2 = .16$ and $F_{1,20} = t_{20}^2 = 4.00$. Note that for these conditions the sampling theorists's .05 critical value of F is $F_{1,20}(.05) = (2.086)^2 = 4.35$ with an associated $R^2 = .18$. Thus the 5% F value is somewhat larger than the Bayesian indifference ($K_{12} = 1$) value of 4.0. Alternatively, an $R^2 = .16$ leads to $K_{12} \doteq 1$ while an $R^2 = .18$ is associated with the sampling theorists's .05 critical value of F. On the other hand, a .01 critical value of F is 8.10, with an associated $R^2 = .29$ which is far from the F value 4.0, or $R^2 = .16$ which yields $K_{12} \doteq 1$.

TABLE 4.2

Values of χ_k^2 and Associated R^2's for $K_{12} \doteq 1$ Using Approximate Formula (3.13) for $\nu_2 = 20$ and $\nu_2 = 100$ and Selected Values for k

	$\nu_2 = 20$		$\nu_2 = 100$		.05 and .01 Critical Values of χ_k^2	
k	χ_k^2	R^{2*}	χ_k^2	R^{2*}	$\chi_k^2(.05)$	$\chi_k^2(.01)$
1	3.00	.13	4.6	.044	3.84	6.63
2	5.99	.23	9.2	.084	5.99	9.21
3	8.99	.31	13.8	.12	7.81	11.30
4	12.0	.37	18.4	.16	9.49	13.30
5	15.0	.43	23.0	.19	11.10	15.10
6	18.0	.47	27.6	.22	12.60	16.80

* Note that $\chi_k^2 = \hat{\beta}'X'X\hat{\beta}/s^2$ and $R^2 = \chi_k^2/(\nu_2 + \chi_k^2)$, where $\nu_2 = n\text{-}k\text{-}1$.

In Table 4.2 values of χ_k^2 and associated values of R^2 which correspond to $K_{12} \doteq 1$ have been tabulated for $\nu_2 = 20$ and $\nu_2 = 100$ and $k = 1,2,\ldots,6$, to gain some information on the quality of the approximation in (3.13). The entries in Table 4.2 have been computed from the large sample approximate formula (3.13), that is $-2\ell n K_{12} \doteq \chi_k^2 - k\ell n\nu_2$. Also shown in Table 4.2 are the .05 and .01 critical values of χ_k^2. For $k = 1$ and $\nu_2 = 20$, the indifference values of χ_1^2 and R^2 are 3.00 and .13, respectively. The latter value can be compared with the more accurate indifference value of $R^2 = .16$ given in Table 4.1. The difference in these values arises because the results in Table 4.2 are based on a cruder approximation than those in Table 4.1. For $\nu_2 = 100$, the corresponding indifference ($K_{12} = 1$) values of R^2 in Tables 4.1 and 4.2 are fairly similar in value. Also, from Table 4.2, the relation of the crude indifference values of χ_k^2 can be compared with the .05 and .01 sampling theory critical values of χ_k^2. For $\nu_2 = 100$, it is seen that for $k = 1$, the indifference value of χ_1^2, namely 4.6 falls between the .05 critical value, 3.84 and the .01 critical value, 6.63. For $k > 2$ and $\nu_2 = 100$, the Bayesian indifference values of χ_k^2 are all larger than the .01 critical values of χ_k^2.

As regards other posterior odds ratios derived in Section III, it is the case that the numerical results in Tables 2.1, 4.1 and 4.2, can be utilized to evaluate them provided that the degrees of freedom and k parameters are suitably reinterpreted. For example, the expressions for K_{AB} in (3.23) and (3.24) can be evaluated if in using the tables, k is replaced by k_2, F_{k,ν_2} is replaced by F_{k_2,ν_2} (note $\nu_2 = \nu_B = n\text{-}k\text{-}1$) and χ_k^2 is replaced by χ_{k2}^2. Similarly, the odds ratios K_{32} and K_{42} in (3.27) and (3.28), respectively can be implemented using the results in the tables by similar redefinitions. Last, the odds ratio, K_{32} in (3.29) can be evaluated from results given in Table 4.1 by use of $K_{32} = K_{32}/K_{42}$, where values of K_{32} and K_{42} can be obtained from entries in Table 4.1. Finally, it is to be noted that in the expression for K_{32} in (3.29), there is an interesting allowance for the possibly differing numbers of parameters in β_1 and β_2.

5. SUMMARY AND CONCLUDING REMARKS

In this paper we have derived approximate posterior odds ratios for sharp null hypotheses which are frequently encountered in regression analyses. These posterior odds ratios are appropriate when little is known regarding parameter values and special attention is given to specific values, e.g. zero values of the regression coefficients. With slight modifications, other special values can be incorporated in the analysis by reparametrizing to convert the null hypotheses to involve zero values. In our work we have employed asymptotic expansions of certain integrals which are very convenient, yield results which can be compared directly with sampling theory analyses, and are quite accurate, as shown in Jeffreys's work. With some extra computational cost, a numerical integration approach, suggested by Dickey (1971) could be applied to obtain slightly more accurate results.

In line with Jeffreys's, Lindley's and some others's previous results, we have found that sampling theorists's usual .05 critical values of test statistics can be far from a Bayesian posterior odds indifference value of one under a variety of circumstances. Whether this finding is interpreted as a systematic flaw in sampling theory practice is of course critically dependent on the nature of the usually implicit loss structure used in sampling theory testing. Cases in which sampling theorists mechanically employ a 5% significance level no matter what the sample size and/or the number of parameters are interpreted as flawed analyses. If sampling theorists and Bayesians carefully consider the underlying loss structure in choosing between or among hypotheses, the above analysis indicates that there can be a compatibility between Bayesian and sampling theory results in testing but, of course their interpretations will differ radically.

While, as pointed out above there can be some degree of compatibility between Bayesian and sampling theory testing results, the direct interpretation of sample evidence, as reflected in F statistics or R^2 values in terms of posterior odds ratios stands in marked contrast to sampling theorists's and others's unclear interpretations of sample evidence in terms of "p-values", and/or values of R^2 or of $\bar{R}^2$, the adjusted coefficient of determination. As mentioned above, a p-value associated with the value of a test statistic is not at all an accurate measure of the posterior probability associated with a null hypothesis. However, it should be noted that most of the posterior odds ratios derived above are monotonically increasing functions of the p-values associated with t or F statistics involved in the posterior odds ratios. Thus there is some rationale for considering p-values; however, since posterior odds ratios have a direct interpretation and explicitly reflect the prior information employed, their use is preferable to the use of p-values. Also, posterior odds ratios on alternative hypotheses can be employed, as described below to average estimates (and predictions) over alternative hypotheses when posterior odds ratios do not yield a clear-cut choice of a particular hypothesis.

In terms of the hypotheses considered above, it is possible to use their associated posterior odds ratios to obtain optimal (relative to quadratic loss) Bayesian "pre-test" point estimates --see Zellner and Vandaele (1974, pp. 640-641). For example with respect to the hypotheses $H_1 : \beta = \mathbf{0}$ and $H_2 : \beta \neq \mathbf{0}$, the point estimate that is optimal relative to quadratic loss is $\ddot{\beta} = P_1\mathbf{0} + (1\text{-}P_1)\tilde{\beta} = (1\text{-}P_1)\tilde{\beta} = (1+K_{12})^{-1}\tilde{\beta}$, where P_1 is the posterior probability for H_1, $K_{12} = P_1/(1\text{-}P_1)$ is the posterior odds ratio for H_1 and H_2, and $\tilde{\beta}$ is the posterior mean for β under H_2. With the prior *pdf* (3.7) which we have employed under H_2, $\tilde{\beta}$ will be close to $\tilde{\beta}$, the least-squares estimate. Thus $\ddot{\beta} \doteq$

$(1+K_{12})^{-1}\hat{\beta}$ where K_{12} given in (3.11) is a function of the usual F statistic. Also note that the "shrinkage factor", $(1+K_{12})^{-1}$ is between zero and one with a value near zero when K_{12} is large and a value near one when K_{12} is small. This shrinkage factor can be compared with others which have appeared in the sampling theory literature —see Zellner and Vandaele (1975, p. 639).

Finally, it would be interesting to compare the posterior odds ratios derived above with others based on more informative prior distributions.

APPENDIX

Herein we evaluate the integrals appearing in equation (3.22) of the text. The integral in the numerator, denoted by I_N is:

$$I_N \propto \int |X_1'X_1|^{1/2}\sigma^{-(n+1)}\exp\{-[\nu_A s_A^2 + n(\alpha\bar{y})^2 + (\eta-\hat{\eta})'X_1'X_1(\eta-\hat{\eta})]/2\sigma^2\}d\alpha d\eta d\sigma. \tag{A.1}$$

where ν_A, s_A^2, $\bar{y}$ and $\hat{\eta}$ have been defined in the text in connection with (3.18). We can integrate over α and the k_1 elements of η using properties of univariate and multivariate normal *pdf*'s, respectively to obtain:

$$I_N \propto (2\pi)^{(k_1+1)/2}n^{-1/2}\int_0^\infty \sigma^{-\nu_A}\exp\{-\nu_A s_A^2/2\sigma^2\}d\sigma. \tag{A.2}$$

Using properties of the inverted gamma *pdf*, the integral in (A.2) can be evaluated to yield:

$$I_N \propto (2\pi)^{(k_1+1)/2}n^{-1/2}\Gamma(\nu_A/2)\,(1/2)(2/\nu_A s_A^2)^{\nu_A/2}. \tag{A.3}$$

The integral in the denominator of (3.22) in the text, denoted by I_D is:

$$I_D \propto \int h(\beta_2|\sigma)|X_1'X_1|^{1/2}\sigma^{-(n+1)}\exp\{-[\nu_B s_B^2 + n(\alpha\text{-}y)^2 + (\eta-\hat{\eta})'X_1'X_1(\eta-\hat{\eta}) + (\beta_2-\hat{\beta}_2)'V'V(\beta_2-\hat{\beta}_2)]/2\sigma^2\}d\alpha d\eta d\sigma \tag{A.4}$$

where $h(\beta_2|\sigma)$ is given in (3.21b) of the text and ν_B, s_B^2, $\hat{\eta}$, $\hat{\beta}_2$ and V have been defined in connection with (3.19). The integration over α and η can be performed exactly using properties of normal distributions to yield:

$$I_D \propto (2\pi)^{(k_1+1)/2}n^{-1/2}\int h(\beta_2|\sigma)\sigma^{-(n-k_1)}\exp\{-[\nu_B s_B^2 + (\beta_2-\hat{\beta}_2)'V'V(\beta_2-\hat{\beta}_2)]/2\sigma^2\}d\beta_2 d\sigma \tag{A.5}$$

The integration over β_2 can be done approximately by noting that is it equivalent to obtaining the expectation of the bounded function $h(\beta_2|\sigma)$. Following Jeffreys's approach and also Cramér's (1946, p. 353) approximation results, we have on integrating approximately with respect to the k_2 elements of β_2

$$I_D \propto (2\pi)^{(k+1)/2} n^{-(k_2+1)/2} c_B \int_0^\infty (1 + \hat{\beta}_2' V' V\hat{\beta}_2/n\sigma^2)^{-(k_2+1)/2}$$
$$\cdot\ \sigma^{-(n-k_1)} \exp\{-\nu_B s_B^2/2\sigma^2\} d\sigma$$

Large values of the second two factors in the integrand of this last expression are near $\nu_B s_B^2/n$. If, as Jeffreys (1967, p. 272) does, we substitute $\sigma^2 = \nu_B s_B^2/n$ in the first, slowly varying factor of the integrand, and then integrate with respect to σ, the result is:

$$I_D \propto c_B 2\pi^{(k+1)/2} n^{-(k_2+1)/2} (1 + \hat{\beta}_2' V' V\hat{\beta}_2/\nu_B s_B^2)^{-(k_2+1)/2} \tag{A.6}$$
$$\times\, \Gamma(\nu_A/2)\,(1/2)\,(2/\nu_B s_B^2)^{\nu_A/2}$$

Then, using (A.3) and (A.6)

$$K_{AB} \doteq \frac{I_N}{I_D} = \frac{1}{c_B}\left(\frac{n}{2\pi}\right)^{k_2/2}\left(\frac{\nu_B s_B^2}{\nu_A s_A^2}\right)^{\nu_A/2}(1+\hat{\beta}_2' V' V\hat{\beta}_2/\nu_B s_B^2)^{(k_2+1)/2} \tag{A.7}$$

Now c_B, the normalizing constant in (3.21b) is

$c_B = \Gamma[(k_2+1)/2]/\pi^{(k_2+1)/2}$ and $\nu_A s_A^2 = \nu_B s_B^2 + \hat{\beta}_2' V' V\hat{\beta}_2$. Thus,

$$K_{AB} \doteq b\,(\nu_B/2)^{k_2/2}\,(\nu_B s_B^2/\nu_A s_A^2)^{(\nu_B-1)/2} \tag{A.8}$$

with $b = \pi^{1/2}/\Gamma[(k_2+1)/2]$ and where in going from (A.7) to (A.8) $(n/2)^{k_2/2}$ has been replaced by $(\nu_B/2)^{k_2/2}$ which to the order of the approximation is equivalent. (A.8) is exactly the expression in (3.23) in the text. Further, on substituting $\nu_B s_B^2/\nu_A s_A^2 = (1-R_B^2)/(1-R_A^2)$ in (A.8), the second line of (3.23) is obtained. Finally, from $\nu_A s_A^2 = \nu_B s_B^2 + \hat{\beta}_2' V' V\hat{\beta}_2$, $\nu_A s_A^2/\nu_B s_B^2 = 1 + \hat{\beta}_2' V' V\hat{\beta}_2/\nu_B s_B^2 = 1 + (k_2/\nu_B)F_{k_2\nu_B}$, which when utilized in (A.8) gives the third line of (3.23).

ACKNOWLEDGEMENTS

Research financed by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, U. of Chicago and by National Science Foundation Grant, SOC 77-15662.

REFERENCES

CRAMER, H. (1946) *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

DEGROOT, M.H. (1973) Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio, *J. Amer. Statist. Assoc.* **68**, 966-969.

DICKEY, J.M. (1971) The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters *Ann. Math. Statist.* **42**, 204-223.

— (1975) Bayesian Alternatives to the F-test and Least-Squares Estimates in the Normal Linear Model. In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, (S.E. Fienberg and A. Zellner eds.), 515-554 Amsterdam: North-Holland.

— (1977) Is the Tail Area Useful as an Approximate Bayes factor? *J. Amer. Statist. Assoc.* **72**, 138-142.

GAVER, K.M. and M.S. GEISEL (1974) Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods, in *Frontiers of Econometrics,* (P. Zarembka ed.) New York: Academic Press.

GEISEL, M.S. (1970) Comparing and Choosing Among Parametric Statistical Models: A Bayesian Analysis with Macroeconomic Application. Ph. D. Thesis. University of Chicago.

JAYNES, E.T. (1976) Confidence Intervals Vs. Bayesian Intervals, in *Foundations of Probability. Theory, Statistical Inference, and Statistical Theories of Science*, (W.L. Harper and C.A. Hooker eds.) 175-213. Dordrecht-Holland: D. Reidel.

JEFFREYS, H. (1957), *Scientific Inference* (2nd ed.) Cambridge: University Press.

— (1967) *Theory of Probability* (3rd rev. ed.), Oxford: University Press.

(1980) "Some General Points in Probability Theory", in *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (A. Zellner ed.) 451-453. Amsterdam: North-Holland.

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

LEMPERS, F.B. (1971), *Posterior Probabilities of Alternative Linear Models,* Rotterdam: University Press.

LINDLEY, D.V. (1957) A Statistical Paradox, *Biometrika* **44**, 187-192.

— (1961) The Use of Prior Probability Distributions in Statistical Inference and Decisions, in *Proc. Fourth Berkeley Symp.* (J. Neyman, ed.) 453-468. Berkeley: University of California Press.

SCHWARZ, G. (1978) Estimating the Dimension of a Model, *Ann. Statist.* **6**, 461-464.

THORNBER, E.H. (1966) Applications of Decision Theory to Econometrics, Ph. D. thesis, University of Chicago.

ZELLNER, A. (1971), *An Introduction to Bayesian Inference in Econometrics,* New York: Wiley

— and W. VANDAELE (1975) Bayes-Stein Estimators for k-means, Regression and Simultaneous Equation Models, in *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* (S.E. Fienberg and A. Zellner eds.) 627-653 Amsterdam: North-Holland.