NONINFORMATIVE PRIORS FOR INFERENCES IN EXPONENTIAL REGRESSION MODELS

by

James O. Berger
Department of Statistics
Purdue University
West Lafayette, IN, USA

and

Keying Ye
Department of Statistics
Purdue University
West Lafayette, IN, USA

Technical Report # 90–05C

Department of Statistics
Purdue University

January 1990

Noninformative Priors for Inferences in Exponential Regression Models*

Keying Ye and James O. Berger

Department of Statistics, Purdue University, West Lafayette, Indiana 47907, U.S.A.

Abstract

In the exponential regression model

$$Y_{ij} = \alpha + \beta\rho^{x+x_i a} + \epsilon_{ij},$$

x and a known, inference concerning ρ is notoriously difficult, even when using the Bayesian noninformative prior approach (cf., Mitchell (1963)). The reference prior approach of Bernardo (1979), as modified by Berger and Bernardo (1989a,b), is considered, and shown to yield very satisfactory inferences. Estimation and credible sets are considered in a specific example from Patterson (1960). The preferred reference prior is thus the first generally recommendable noninformative prior for the problem.

*Research supported by the National Science Foundation, Grant DMS-8702620

1 Introduction

Consider the exponential regression model given by

$$Y_{ij} \sim N(\alpha + \beta\rho^{x+x_i a}, \sigma^2), \tag{1}$$

where $\alpha, \beta \in R, 0 < \rho < 1$ and $x \geq 0, a > 0$, x and a known constants, $0 \leq i \leq k-1, 1 \leq j \leq m$, the x_i's are known nonnegative regressors with $x_i \neq x_j$ for $i \neq j$ and the variance $\sigma^2 > 0$ is an unknown constant. Without loss of generality, it is assumed that $x_i < x_j$ for $i < j$.

Bayesian inference for ρ is desired, based on a noninformative prior. Mitchell (1963, 1967) considered the above problem for a variety of possible noninformative priors. Common adhoc noninformative priors, such as $\pi(\alpha, \beta, \sigma, \rho) \propto \sigma^{-1}$, were shown to yield improper posterior densities for ρ, invalidating their use in Bayesian inference (see also Cox & Hinkley, 1974). Faced with this surprising phenomenon, Mitchell considered various adhoc modifications of the noninformative priors, modifications designed to produce proper posterior densities for ρ. Properties of these modifications and ensuing Bayesian inferences were then studied (see Section 4 for discussion).

Consonni and Veronese (1989) also considered the model of Mitchell and analyzed use of $\pi(\alpha, \beta, \sigma, \rho) \propto \sigma^{-1}$ using a finitely-additive approach. They obtained a proper (but only finitely additive) posterior distribution for ρ. Though proper, this posterior distribution is of limited practical use, since it only has mass (or adherent mass) at $\rho = 0$ and/or $\rho = 1$.

Our interest in this study arose out of a desire to "test" the reference prior algorithm, initiated by Bernardo (1979) and further developed by Berger and Bernardo (1989a,b), for development of a noninformative prior on this challenging and troublesome model of considerable practical importance. The hope was that this "automatic" reference prior approach would yield a prior giving a proper posterior with good properties. This would not only provide additional evidence in support of the reference prior algorithm, but might yield a generally acceptable noninformative prior for the nonlinear regression model. In Section 4, evidence of the success of the reference prior will be presented.

The most well-known "automatic" method of generating a noninformative prior is that of Jeffreys

(1961). Mitchell (1963) determined the Jeffreys prior for this problem, but had several criticisms of it. One criticism was that it depended on β in an unnatural way. A second criticism was that it depended on the covariates $\{x_i\}$ and number of treatments k, seemingly unnatural for a prior distribution. We share the concern about β (see Section 2), but not that of dependence on the $\{x_i\}$ and k. It has become well accepted that noninformative priors must depend on the statistical model, which here includes the $\{x_i\}$ and k. (See Bernardo, 1979, and Berger, 1985, for discussion.) Dependence on covariates is not particularly attractive (especially if they, themselves, are random), but is arguably a necessary evil if one desires a sound noninformative prior. We stress this because the reference priors will also be seen to depend on the $\{x_i\}$ and k.

Section 2 contains the development of the reference priors, and also gives the Jeffreys prior as determined by Mitchell. In Section 3, the marginal posterior densities of ρ will be determined for the reference and Jeffreys priors. Part of this development is a demonstration that the reference and Jeffreys priors yield proper posterior densities for ρ. Also of considerable interest is that the reference posteriors are much simpler to work with than the Jeffreys posterior, precisely because the reference priors do not have the unnatural dependence on β that is found in the Jeffreys prior.

In Section 4, comparisons of the inferences under the various priors, and with maximum likelihood estimators, will be undertaken in an example. Estimates and their associated standard errors will be considered, along with actual frequentist coverage of the implied confidence procedures (a common method of evaluating noninformative priors – cf. Berger & Bernardo, 1989b, Efron, 1986, Stein, 1985). Section 5 contains discussion. Most of the technical details are put in Appendices.

2 The reference priors and the Jeffreys prior

2.1 Preliminaries

The likelihood function for $\alpha, \beta, \sigma, \rho$ in the exponential regression model is

$$f(\mathbf{y}|\alpha,\beta,\sigma,\rho) \propto \sigma^{-mk} \exp\{-\frac{1}{2\sigma^2}[s^2 + m\sum_{i=0}^{k-1}(\overline{y}_{i.} - \alpha - \beta\rho^{x+x_i a})^2]\}, \tag{2}$$

where $\mathbf{y}$ is a vector consisting of all the observations, $\overline{y}_{i.} = \frac{1}{m}\sum_{j=1}^{m} y_{ij}$, $\overline{y} = \frac{1}{k}\sum_{i=0}^{k-1}\overline{y}_{i.}$, and $s^2 = \sum_{i=0}^{k-1}\sum_{j=1}^{m}(y_{ij}-\overline{y}_{i.})^2 = s_{yy}^2 - s_y^2$. Here $s_{yy}^2 = \sum_{i=0}^{k-1}\sum_{j=1}^{m}(y_{ij}-\overline{y})^2$ and $s_y^2 = m\sum_{i=0}^{k-1}(\overline{y}_{i.}-\overline{y})^2$.

The Fisher information matrix is

$$H(\rho,\alpha,\beta,\sigma) = \frac{m}{\sigma^2}\begin{pmatrix} H_1(\rho,\beta) & \mathbf{0} \\ \mathbf{0}^t & 2k \end{pmatrix}, \tag{3}$$

where

$$H_1(\rho,\beta) = \begin{pmatrix} \beta^2\rho^{2(x-1)}\sum_{i=0}^{k-1}(x+x_ia)^2\rho^{2x_ia} & \beta\rho^{x-1}\sum_{i=0}^{k-1}(x+x_ia)\rho^{x_ia} & \beta\rho^{2x-1}\sum_{i=0}^{k-1}(x+x_ia)\rho^{2x_ia} \\ \beta\rho^{x-1}\sum_{i=0}^{k-1}(x+x_ia)\rho^{x_ia} & k & \rho^x\sum_{i=0}^{k-1}\rho^{x_ia} \\ \beta\rho^{2x-1}\sum_{i=0}^{k-1}(x+x_ia)\rho^{2x_ia} & \rho^x\sum_{i=0}^{k-1}\rho^{x_ia} & \rho^{2x}\sum_{i=0}^{k-1}\rho^{2x_ia} \end{pmatrix}.$$

When $k \leq 2$, it can be shown that the information matrix (3) is singular, which precludes application of the reference prior or the Jeffreys prior approaches. Thus we henceforth assume that $k \geq 3$.

For later use, several functions are defined as follows:

$$\begin{aligned} p_1(w) &= (\sum_{i=0}^{k-1} w^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1} w^{x_i})^2)(\sum_{i=0}^{k-1} x_i^2 w^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1} x_i w^{x_i})^2) \\ &\quad -(\sum_{i=0}^{k-1} x_i w^{2x_i} - \frac{1}{k}\sum_{i=0}^{k-1} w^{x_i}\sum_{i=0}^{k-1} x_i w^{x_i})^2, \end{aligned} \tag{4}$$

$$p_2(w) = \sum_{i=0}^{k-1} w^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1} w^{x_i})^2, \tag{5}$$

$$q = \sum_{i=0}^{k-1} x_i^2 - \frac{1}{k}(\sum_{i=0}^{k-1} x_i)^2, \tag{6}$$

and

$$Q = \sum_{i=0}^{k-1} x_i^2 \sum_{i=0}^{k-1} x_i^4 - (\sum_{i=0}^{k-1} x_i^3)^2 - \frac{1}{k}(\sum_{i=0}^{k-1} x_i^2)^3 - \frac{1}{k}(\sum_{i=0}^{k-1} x_i)^2 \sum_{i=0}^{k-1} x_i^4 + \frac{2}{k}\sum_{i=0}^{k-1} x_i \sum_{i=0}^{k-1} x_i^2 \sum_{i=0}^{k-1} x_i^3. \tag{7}$$

2.2 Reference priors

In this problem, we are considering ρ as the parameter of interest and α, β, σ as nuisance parameters. Bernardo (1979) suggests determining the reference prior by a two step procedure in which, first, one finds the conditional reference prior for the nuisance parameters given the parameter of interest; and then one finds the reference prior for the parameter of interest in the marginal model formed by integrating out the nuisance parameter with respect to the conditional reference prior. Berger and Bernardo (1989a,b,c) propose allowing multiple groups "ordered" in terms of inferential importance, with the reference prior being determined through a succession of analyses for the implied conditional problems. (Note that this reference prior algorithm differs from consideration of "independent" Jeffreys priors, as done by Mitchell, 1963.) Berger and Bernardo particularly recommend the reference prior based on having each parameter in its own group, i.e., having each conditional reference prior be only 1-dimensional.

In this subsection, all the reference priors for the various group orderings of $\rho, \alpha, \beta, \sigma$ will be given. Notation such as $\{\rho, \alpha, (\beta, \sigma)\}$ will be used to specify the groups and the importance of the parameters; $\{\rho, \alpha, (\beta, \sigma)\}$ means there are 3 groups, with ρ being most important and the group (β, σ) being least important. To compute the reference prior for this group ordering (for instance), we thus first derive the conditional prior $\pi(\beta, \sigma|\rho, \alpha)$, use it to determine a conditional reference prior $\pi(\alpha|\rho)$ in a marginalized model, and use both of them to derive a marginal reference prior $\pi(\rho)$.

Theorem 2.1 *For the exponential regression model as described in Section 1 with $k \geq 3$, the reference priors for the various possible group orderings are as follows:*

Group Ordering	Reference Prior
$\{\rho, \alpha, \beta, \sigma\}$ *or* $\{\rho, (\alpha, \beta), \sigma\}$ *and with all permutations of* α, β, σ	$\frac{1}{\sigma}\rho^{x-1}p(\rho^a)$
$\{\rho, \alpha, (\beta, \sigma)\}$ *and with all permutations of* α, β, σ	$\frac{1}{\sigma^2}\rho^{x-1}p(\rho^a)$
$\{\rho, (\alpha, \beta, \sigma)\}$ *and with all permutations of* α, β, σ	$\frac{1}{\sigma^3}\rho^{x-1}p(\rho^a)$

where (see (4) and (5)) $p(\rho^a) = p_1(\rho^a)/p_2(\rho^a)$.

Proof: See Appendix A. □

Note that the reference priors only depend on σ and ρ. The only differences for the different group orderings are in the power of σ^{-1}. To verify that the reference priors yield proper posteriors, we must discuss the behavior of the function $p(\rho)$ for $\rho \in (0,1)$. Since $p_1(\rho)$ and $p_2(\rho)$ are continuous, nonnegative and bounded functions, and $p_2(\rho)$ is non-zero on $(0,1)$ when the x_i's are not all equal, the integrability of $p(\rho)$ on $(0,1)$ is determined by its behavior near 0 and 1.

In Appendix B, the following results are established (see (33) , (34), (35) and (36) in Appendix B):

(a) For $\rho \sim 0$:

$$p_1^2(\rho) = (1 - \frac{2}{k})(x_1 - x_0)^2 \rho^{2(x_0+x_1)} + o(\rho^{2(x_0+x_1)}), \tag{8}$$

and

$$p^2(\rho) = \frac{k-2}{k-1}(x_1 - x_0)^2 \rho^{2x_1} + o(\rho^{2x_1}). \tag{9}$$

(b) For $\rho \sim 1$:

$$p_1^2(\rho) = \frac{Q}{4}(1-\rho)^4 + o((1-\rho)^4), \tag{10}$$

and

$$p^2(\rho) = \frac{Q}{4q}(1-\rho)^2 + o((1-\rho)^2), \tag{11}$$

where q and Q are defined in (6) and (7). Both q and Q are positive when the x_i's are not all equal.

From (9) and (11), it can be concluded that all reference priors for ρ behave near 0 and 1 as

$$\begin{aligned} \pi_s^r(\rho,\alpha,\beta,\sigma) &\propto \frac{\rho^{x-1}}{\sigma^s} p(\rho^a) \\ &\sim \begin{cases} \frac{1}{\sigma^s}\sqrt{\frac{k-2}{k-1}}(x_1 - x_0)\rho^{ax_1+x-1} & \rho \sim 0 \\ \frac{1}{2\sigma^s}\sqrt{\frac{Q}{q}}(1-\rho^a) & \rho \sim 1 \end{cases}, \end{aligned} \tag{12}$$

where s is $1,2$ or 3. This will be used in Section 3 to show that the marginal reference posteriors for ρ are proper.

2.3 The Jeffreys prior

In this problem, the Jeffreys prior was first established in Mitchell (1963). In general, the Jeffreys prior is the square root of the determinant of the Fisher information matrix. From (30) and (31) in Appendix A, the following result is immediate:

Theorem 2.2 *For the exponential regression model as described in Section 1 with $k \geq 3$, the Jeffreys prior has the form*

$$\pi^J(\rho, \alpha, \beta, \sigma) \quad \propto \quad \frac{|\beta|\rho^{2x-1}}{\sigma^4} p_1(\rho^a). \tag{13}$$

Note that the Jeffreys prior depends on ρ, σ, and also β, while the reference priors given in section (2.2) depend only on ρ and σ. This causes considerable complications in computation with the Jeffreys prior since the integration over β cannot be done in closed form here, as it can for the reference priors. Furthermore, Mitchell (1963) argued against the presence of $|\beta|$ in (13), on intuitive grounds. We will discuss this later in Section 4.

By equations (8) and (10), the limiting behavior of the Jeffreys prior near 0 and 1 is:

$$\pi^J(\rho, \alpha, \beta, \sigma) \sim \begin{cases} \frac{\sqrt{(1-\frac{2}{k})}}{\sigma^4} |\beta| \rho^{a(x_0+x_1)+2x-1} & \rho \sim 0 \\ \frac{1}{2\sigma^4}\sqrt{Q}|\beta|(1-\rho^a)^2 & \rho \sim 1 \end{cases}. \tag{14}$$

This, also, will be used to establish the propriety of the associated marginal posterior for ρ.

3 Marginal posterior densities of ρ

3.1 Marginal reference posterior densities of ρ

Mitchell (1963) computed the posterior density of ρ, for $x_i = i$, $0 \leq i \leq k-1$, using the prior σ^{-1} and integrating out with respect to α, β, σ, obtaining

$$\pi(\rho|\mathbf{y}) \propto \{\rho^x(1-\rho^a)h(\rho;\mathbf{y})\}^{-1}, \tag{15}$$

where $h(\rho;\mathbf{y})$ is a continuous and bounded function on $(0,1)$. Unfortunately, this turns out to be improper. Via similar calculation, the posterior densities of ρ using reference priors and the Jeffreys prior can be found. The following two conditions will be assumed:

(a) *there do not exist* $\overline{\rho} \in (0,1)$ *and constants* C_1, C_2 *such that* $y_{ij} \equiv C_1 + C_2\overline{\rho}^{x_i}$ *for all* i *and* j;

(b) *there do not exist constants* C_1 *and* C_2 *such that* $y_{ij} \equiv C_1 + C_2 x_i$, *for all* i *and* j. (16)

Note that (a) excludes the case in which the data exactly fits the given model (residuals equal to zero) and (b) excludes the case in which the data exactly fits a linear regression model. Neither case can be handled with improper priors. Note, of course, that these cases have probability zero of occurring under the given model.

Theorem 3.1 *For the exponential regression model described in Section 1 with* $k \geq 3$, *suppose that (a) and (b) of (16) are satisfied. Then, for the reference priors* $\pi_s^r(\rho, \alpha, \beta, \sigma) = \sigma^{-s}\rho^{x-1}p(\rho^a)$ *given in Section 2.2, the marginal posterior densities of* ρ *are proper and are given by*

$$\pi_s^r(\rho|\mathbf{y}) = \frac{K(s)p(\rho^a)}{\rho^{1+ax_0}(1-\rho^a)h(\rho^a; s, \mathbf{y})}, \tag{17}$$

where $K(s)$ *are the normalization constants for* $s = 1, 2$ *or* 3. *Here*

$$h(\rho; s, \mathbf{y}) = \{[s_{yy}^2 - m d_k^2(\rho; \mathbf{y})/p_2^2(\rho)]^{km+s-3}\frac{p_2^2(\rho)}{\rho^{2x_0}(1-\rho)^2}\}^{\frac{1}{2}}, \tag{18}$$

where

$$d_k(\rho; \mathbf{y}) = \sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y})\rho^{x_i}. \tag{19}$$

Proof: The form (17) follows directly from calculation using the reference priors obtained in Theorem 2.2 and the likelihood function given in (2).

To show that these densities are proper, observe that $h(\rho; s, \mathbf{y})$ is continuous and bounded for $\rho \in (0,1)$ when s and $\mathbf{y}$ are given. We need to consider the behavior of $h^{-1}(\rho; s, \mathbf{y})$ on $(0,1)$. Since, by the Schwartz inequality,

$$\begin{aligned} d_k^2(\rho; \mathbf{y}) &= [\sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y})\rho^{x_i}]^2 = [\sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y})(\rho^{x_i} - \frac{1}{k}\sum_{i=1}^{k-1}\rho^{x_i})]^2 \\ &\leq \sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y})^2 \sum_{i=0}^{k-1}(\rho^{x_i} - \frac{1}{k}\sum_{i=0}^{k-1}\rho^{x_i})^2 = p_2^2(\rho)\sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y})^2, \end{aligned}$$

it follows that

$$s_{yy}^2 - md_k^2(\rho;\mathbf{y})/p_2^2(\rho) \geq s^2 + s_y^2 - m\sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y})^2 = s^2. \tag{20}$$

Therefore, $h(\rho; s, \mathbf{y}) = 0$ for some $\rho \in (0,1)$ if and only if equality holds in (20) for some ρ and $s^2 = 0$. Equality holds in (20) for some ρ if and only if there is a $\overline{\rho} \in (0,1)$ such that $\overline{y}_{i.} - \overline{y} = K(\overline{\rho}^{x_i} - \frac{1}{k}\sum_{i=0}^{k-1}\overline{\rho}^{x_i})$, $0 \leq i \leq k-1$ for some constant K. If this were to hold and $s^2 = 0$, then condition (a) would be violated. Thus, $h^{-1}(\rho; s, \mathbf{y}) > 0$ is well defined for $\rho \in (0,1)$.

To see the limiting behavior of $h(\rho; s, \mathbf{y})$ at the endpoints, observe that

$$h(0_+; s, \mathbf{y}) = \lim_{\rho \to 0^+} h(\rho; s, \mathbf{y}) = [(1 - \frac{1}{k})]^{\frac{1}{2}}[s_{yy}^2 - \frac{km(\overline{y}_{0.} - \overline{y})^2}{(k-1)}]^{\frac{km+s-3}{2}}, \tag{21}$$

and

$$h(1_-; s, \mathbf{y}) = \lim_{\rho \to 1^-} h(\rho; s, \mathbf{y}) = q^{\frac{1}{2}}(k; \mathbf{x})\{s_{yy}^2 - [\sum_{i=0}^{k-1} x_i(\overline{y}_{i.} - \overline{y})]^2/q\}^{\frac{km+s-3}{2}}. \tag{22}$$

By (21), since

$$\begin{aligned}
(k-1)s_{yy}^2 - km(\overline{y}_{0.} - \overline{y})^2 &= (k-1)s^2 + (k-1)m\sum_{i=1}^{k-1}(\overline{y}_{i.} - \overline{y})^2 - km(\overline{y}_{0.} - \overline{y})^2 \\
&\geq m[(k-1)\sum_{i=1}^{k-1}(\overline{y}_{i.} - \overline{y})^2 - (\overline{y}_{0.} - \overline{y})^2] \\
&\geq m[(\sum_{i=1}^{k-1}(\overline{y}_{i.} - \overline{y}))^2 - (\overline{y}_{0.} - \overline{y})^2] = 0,
\end{aligned}$$

then $h(0_+; s, \mathbf{y}) \geq 0$. Equality holds if and only if all $\overline{y}_{i.}$ are equal and $y_{ij} = \overline{y}_{i.}$ for all i and j, which is excluded under condition (a).

By (22), a similar technique shows that $h(1_-; s, \mathbf{y}) \geq 0$, with equality holding if and only if $y_{ij} = \overline{y}_{i.}$ for all i, j, and $\overline{y}_{i.} = C_1 + C_2 i$, where C_1 and C_2 are constants. This latter case is excluded under condition (b).

Therefore, $h^{-1}(\rho; s, \mathbf{y})$ is uniformly bounded on $(0,1)$. Using (9) and (11), it can be concluded that the limiting behavior of the marginal reference posterior densities of ρ near either 0 or 1 is

$$\pi_s^r(\rho|\mathbf{y}) \sim \begin{cases} K(s)\sqrt{\frac{(k-2)}{k-1}}(x_1 - x_0)\rho^{a(x_1-x_0)-1}/h(0_+; s, \mathbf{y}) & \rho \sim 0 \\ \frac{K(s)}{2}\sqrt{\frac{Q}{q}}/h(1_-; s, \mathbf{y}) & \rho \sim 1 \end{cases}.$$

Since $a > 0$, it follows that the reference marginal posterior densities are proper on $(0,1)$. □

3.2 Marginal posterior density of ρ using the Jeffreys prior

Rewriting the likelihood function (2) as

$$\begin{aligned} f(\mathbf{y}|\alpha,\beta,\sigma,\rho) &\propto \sigma^{-km}\exp\{-\frac{1}{2\sigma^2}[A(\rho^a,\mathbf{y})+km(\alpha-B(\rho,\beta,\mathbf{y}))^2 \\ &\quad +m\rho^{2x}p_2^2(\rho^a)(\beta-\frac{C(\rho^a,\mathbf{y})}{\rho^x})^2]\}, \end{aligned} \tag{23}$$

where

$$\begin{aligned} A(\rho,\mathbf{y}) &= s_{yy}^2 - md_k^2(\rho,\mathbf{y})/p_2^2(\rho), \\ B(\rho,\beta,\mathbf{y}) &= \overline{y}-\frac{\beta}{k}\rho^x\sum_{i=0}^{k-1}\rho^{x_i a}, \\ C(\rho,\mathbf{y}) &= \frac{d_k(\rho,\mathbf{y})}{\rho^x p_2^2(\rho)}, \end{aligned} \tag{24}$$

we have the following result:

Theorem 3.2 *For the exponential regression model described in Section 1 with $k \geq 3$, suppose that conditions (a) and (b) of (16) are satisfied. Then the marginal posterior density of ρ for the Jeffreys prior is proper and has the form*

$$\pi^J(\rho|\mathbf{y}) = K\frac{p(\rho^a)}{\rho p_2(\rho^a)}[\frac{1}{km}(s_{yy}^2)^{-\frac{km}{2}} + \frac{(s_{yy}^2 - A(\rho^a,\mathbf{y}))^{\frac{1}{2}}}{A(\rho^a,\mathbf{y})^{\frac{1+km}{2}}}\int_0^{(\frac{s_{yy}^2}{A(\rho^a,\mathbf{y})}-1)^{\frac{1}{2}}}\frac{d\gamma}{(1+\gamma^2)^{1+\frac{km}{2}}}], \tag{25}$$

where K is the normalization constant.

Proof: From (13), it follows that the marginal posterior density of ρ is

$$\begin{aligned} \pi^J(\rho|\mathbf{y}) &\propto \rho^{2x-1}p_1(\rho^a)\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty|\beta|\sigma^{-4}f(\mathbf{y}|\alpha,\beta,\sigma,\rho)d\alpha d\beta d\sigma \\ &\propto \rho^{-1}p_1(\rho^a)\int_0^\infty\int_{-\infty}^\infty|\beta|\sigma^{-(4+km)}\exp\{-\frac{1}{2\sigma^2}[A(\rho^a,\mathbf{y})+mp_2^2(\rho^a)[\beta-C(\rho^a,\mathbf{y})]^2]\} \\ &\quad \int_{-\infty}^\infty\exp\{\frac{km}{2\sigma^2}[\alpha-B(\rho,\beta,\mathbf{y})]^2\}d\alpha d\beta d\sigma \\ &\propto \frac{p(\rho^a)}{\rho p_2(\rho^a)}\int_{-\infty}^\infty\int_0^\infty\sigma^{-(3+km)}|\beta+(\frac{s_{yy}^2}{A(\rho^a,\mathbf{y})}-1)^{\frac{1}{2}}|\exp\{-\frac{1}{2\sigma^2}[A(\rho^a,\mathbf{y})+\beta^2]\}d\sigma d\beta \\ &\propto \frac{p(\rho^a)}{\rho p_2(\rho^a)}\int_{-\infty}^\infty\frac{|\beta+(\frac{s_{yy}^2}{A(\rho^a,\mathbf{y})}-1)^{\frac{1}{2}}|}{[A(\rho^a,\mathbf{y})+\beta^2]^{\frac{km+2}{2}}}d\beta \\ &\propto \frac{p(\rho^a)}{\rho p_2(\rho^a)}[\frac{1}{km}(s_{yy}^2)^{-\frac{km}{2}} + \frac{(s_{yy}^2-A(\rho^a,\mathbf{y}))^{\frac{1}{2}}}{A(\rho^a,\mathbf{y})^{\frac{1+km}{2}}}\int_0^{(\frac{s_{yy}^2}{A(\rho^a,\mathbf{y})}-1)^{\frac{1}{2}}}\frac{d\gamma}{(1+\gamma^2)^{1+\frac{km}{2}}}]. \end{aligned}$$

Since $h(\rho; s, \mathbf{y}) = [A(\rho, \mathbf{y})^{km+s-3} \frac{p_2^2(\rho)}{\rho^{2x_0}(1-\rho)^2}]^{\frac{1}{2}}$ (by (18)) and $\frac{p_2^2(\rho)}{\rho^{2x_0}(1-\rho)^2}$ is bounded on $[0,1]$ (by (33) and (35)), the discussion of $h^{-1}(\rho; s, \mathbf{y})$ in the proof of Theorem 3.1 shows that $A^{-1}(\rho, \mathbf{y})$ is uniformly bounded for ρ in $(0,1)$, excluding the cases (a) and (b). Also, $1/s_{yy}^2$ is bounded under condition (a). By (33) and (35) of Appendix B and (9) and (11) in Section 2.2 it follows that

$$\frac{p(\rho^a)}{\rho p_2(\rho^a)} \sim \begin{cases} \sqrt{k(k-2)} \frac{x_1 - x_0}{k-1} \rho^{a(x_1 - x_0)-1} & \rho \sim 0 \\ \frac{\sqrt{Q}}{2q} & \rho \sim 1 \end{cases} .$$

Thus, the marginal posterior density of ρ using the Jeffreys prior is proper. □

4 Comparison in an application

As in Mitchell (1963), we consider the example given by Patterson (1960) concerning a group of data resulting from seven experiments yielding seven different sets of plant yields. Patterson (1960) computed the least-squares estimates of ρ for each experiment using the exponential regression model. Corresponding to the model described in Section 1, the seven experiments had $k = 4$, $m = 1$, $x = 0$, $a = 1$ and $x_i = i$ for $0 \le i \le 3$. In this example, the response variable used is Barley grain and the regressors are the levels of nitrogen used for plants. Here α is the average yield of plants, $-\beta$ measures the maximum change in the yields and ρ represents the efficiency of the fertilizer. Table 1 gives the data.

Level of	*Data Set*						
nitrogen	1	2	3	4	5	6	7
0	29.0	28.9	26.6	32.2	13.5	29.7	15.8
1	35.0	36.1	32.1	38.6	17.1	34.3	23.4
2	37.6	37.5	34.2	38.6	18.7	36.7	28.5
3	37.7	37.2	34.6	39.6	19.1	37.2	32.2

Table 1: Experimental data

For this experimental setup, the reference priors are (see (12) and (13))

$$\pi_s^r(\rho, \alpha, \beta, \sigma) \propto \frac{\psi^r(\rho)}{\sigma^s}, \tag{26}$$

for $s = 1, 2$ or 3, and the Jeffreys prior is

$$\pi^J(\rho, \alpha, \beta, \sigma) \propto \frac{|\beta|\psi^J(\rho)}{\sigma^4}, \tag{27}$$

where

$$\psi^r(\rho) = K(r)(1-\rho)\sqrt{\frac{1+2\rho+4\rho^2+2\rho^3+\rho^4}{3+4\rho+6\rho^2+4\rho^3+3\rho^4}}, \tag{28}$$

and

$$\psi^J(\rho) = K(J)(1-\rho)^2\sqrt{1+2\rho+4\rho^2+2\rho^3+\rho^4}; \tag{29}$$

it happens that ψ^r and ψ^J, the factors of the priors that depend on ρ, have finite integrals over $(0, 1)$, so that it is convenient for comparison to let $K(r)$ and $K(J)$ be the normalization constants such that the integrals of ψ^r and ψ^J over $(0, 1)$ are equal to 1. Figure 1 gives the graphs of ψ^r and ψ^J for this example. Observe that both decrease to zero as $\rho \to 1$. This is what yields propriety of the resulting marginal posteriors for ρ.

For the 7 data sets, the marginal posterior densities for ρ, using the reference priors and the Jeffreys prior, are given in Figures 2-8. Table 2 presents the corresponding marginal posterior means

Data set	*MLE*	*Posterior Means*				*Posterior Modes*			
		Jeffreys	*Reference*			*Jeffreys*	*Reference*		
			s=3	s=2	s=1		s=3	s=2	s=1
1	0.3329	0.3341	0.3352	0.3389	0.3502	0.3321	0.3319	0.3317	0.3308
2	0.1482	0.1489	0.1504	0.1562	0.1800	0.1479	0.1479	0.1478	0.1477
3	0.3375	0.3375	0.3379	0.3390	0.3451	0.3373	0.3373	0.3372	0.3370
4	0.0948	0.1181	0.1320	0.1579	0.2092	0.0933	0.0930	0.0928	0.0913
5	0.3926	0.3929	0.3931	0.3940	0.3977	0.3924	0.3923	0.3922	0.3921
6	0.4267	0.4275	0.4280	0.4293	0.4319	0.4261	0.4259	0.4255	0.4252
7	0.6919	0.6920	0.6920	0.6920	0.6888	0.6919	0.6919	0.6919	0.6918

Table 2: Posterior means and modes and the MLE in estimation of ρ

FIG. 1: COMPARISON OF PSI(R) and PSI(J)

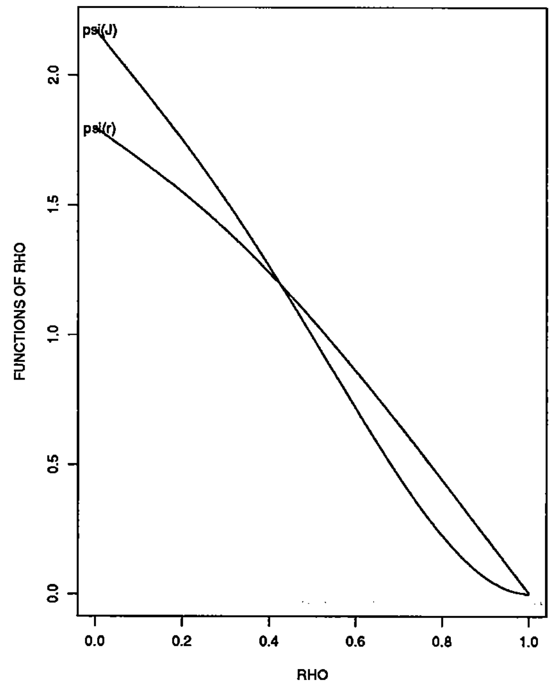


FIG. 2: MARGINAL POSTERIOR FOR RHO FOR DATA SET 1

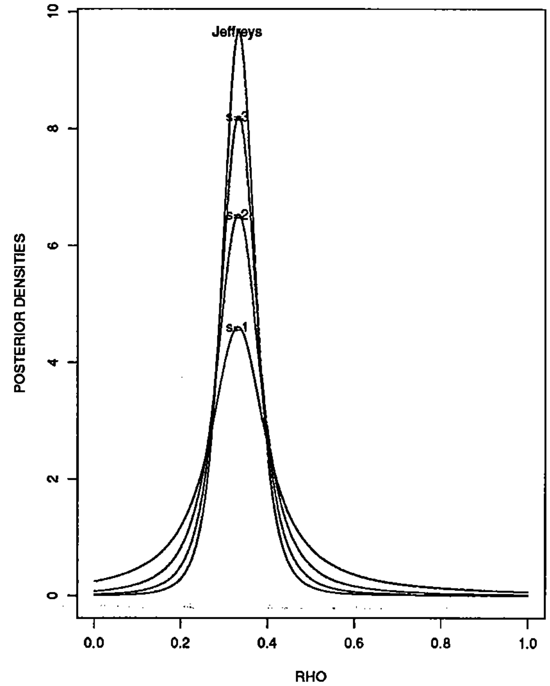


FIG. 3: MARGINAL POSTERIOR FOR RHO FOR DATA SET 2

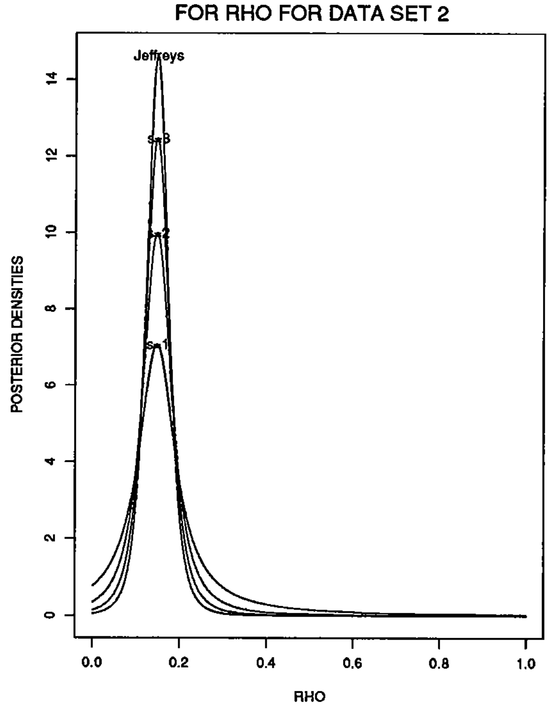


FIG. 4: MARGINAL POSTERIOR FOR RHO FOR DATA SET 3

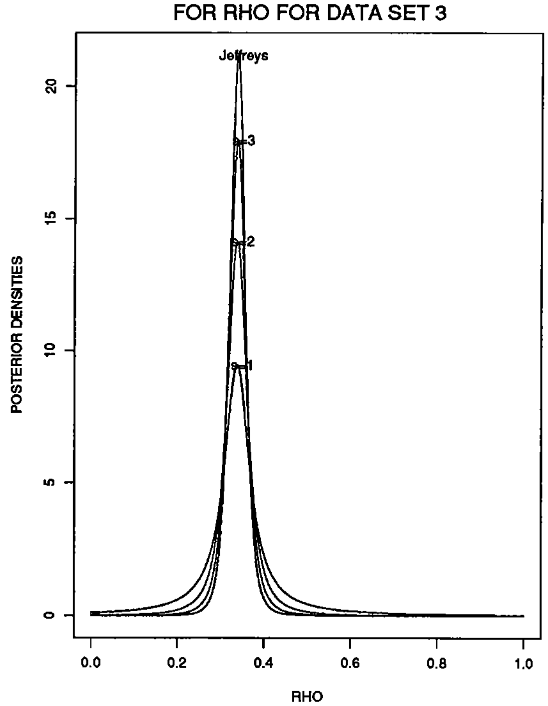


FIG. 5: MARGINAL POSTERIOR
FOR RHO FOR DATA SET 4

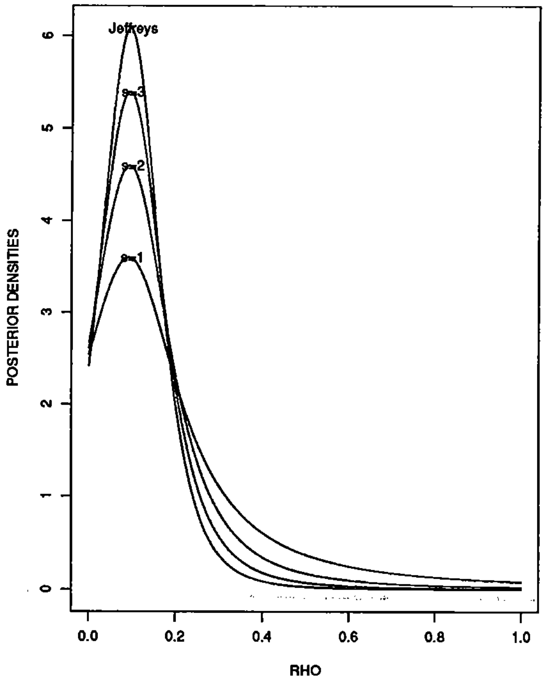


FIG. 6: MARGINAL POSTERIOR
FOR RHO FOR DATA SET 5

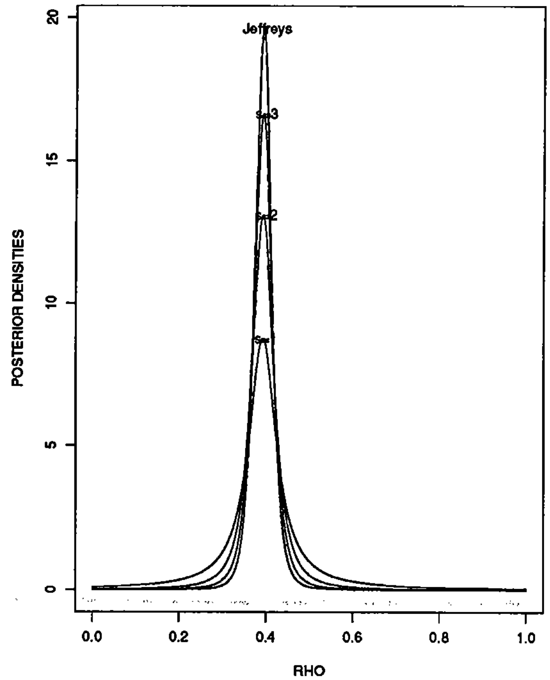


FIG. 7: MARGINAL POSTERIOR
FOR RHO FOR DATA SET 6

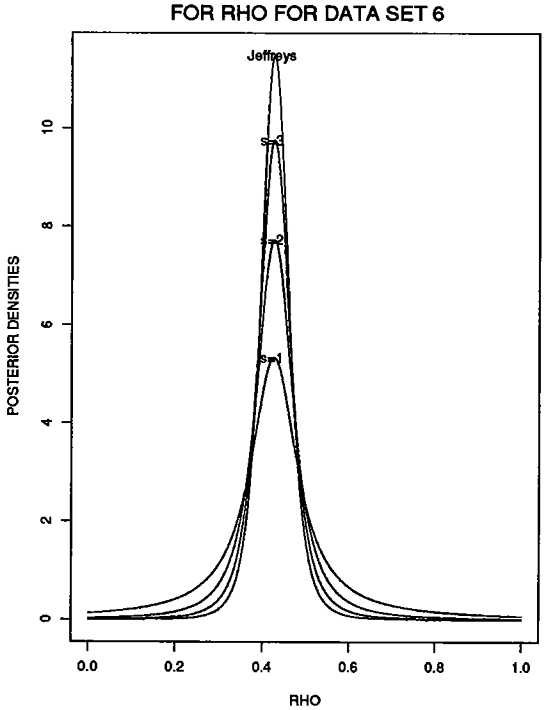
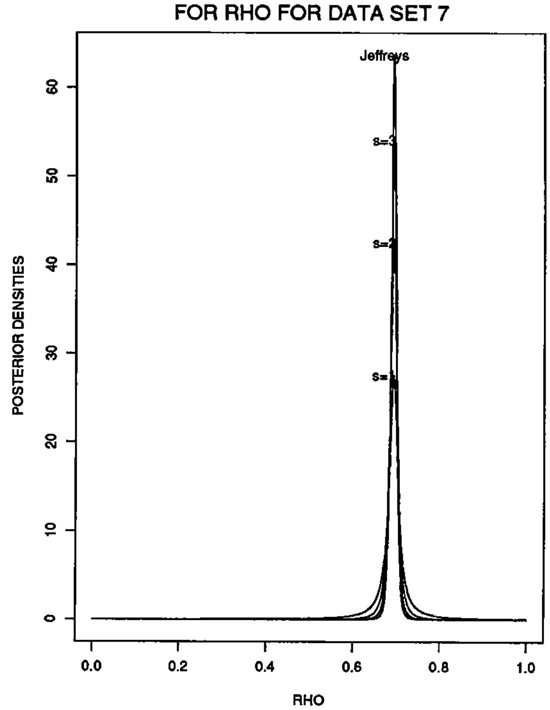


FIG. 8: MARGINAL POSTERIOR
FOR RHO FOR DATA SET 7



Data Set	for MLE(1)	for MLE(2)	for Jeffreys	for reference analyses		
				s=3	s=2	s=1
1	1.5078×10^{-3}	1.6342×10^{-3}	2.9545×10^{-3}	4.9943×10^{-3}	9.8088×10^{-3}	2.1473×10^{-2}
2	6.5984×10^{-4}	7.0867×10^{-4}	1.2440×10^{-3}	2.1340×10^{-3}	4.8844×10^{-3}	1.5305×10^{-2}
3	3.1371×10^{-4}	3.2536×10^{-4}	6.2673×10^{-4}	1.1609×10^{-3}	3.0064×10^{-3}	1.0603×10^{-2}
4	4.9511×10^{-3}	4.0045×10^{-3}	6.3904×10^{-3}	1.0319×10^{-2}	1.8657×10^{-2}	3.5733×10^{-2}
5	3.6530×10^{-4}	3.7876×10^{-4}	7.3010×10^{-4}	1.3465×10^{-3}	3.4070×10^{-3}	1.1373×10^{-2}
6	1.0646×10^{-3}	1.1289×10^{-3}	2.1092×10^{-3}	3.6665×10^{-3}	7.6823×10^{-3}	1.8436×10^{-2}
7	3.4594×10^{-5}	3.4339×10^{-5}	6.9138×10^{-5}	1.3338×10^{-4}	4.5920×10^{-4}	3.6878×10^{-3}

Table 3. Posterior variances and estimated asymptotic variances of the MLE.

and modes for ρ, as well as the MLE; Table 3 presents the corresponding posterior variances for ρ, as well as the estimated asymptotic variances of the MLE for ρ, calculated by two methods (called MLE(1) and MLE(2); see Appendix D). (Table 2 and Table 3 were calculated using the subroutines DCADRE, DBLIN and ZPOLR of IMSL on a VAX11/780 computer; all numbers in the two tables are accurate through the given digits.)

From Table 2, it is clear that the posterior modes are all very similar. The posterior means for the reference posteriors are shifted right, quite substantially for data set 2 and especially for data set 4. The reason is clear from looking at Figures 3 and 5, which exhibit definite skewness. The posterior means and modes for the Jeffreys prior are closest to the MLE, followed by those for the reference priors with $s = 3$, $s = 2$ and $s = 1$, in that order.

From the graphs, it is clear that all posterior densities have the same shape, but with different variances. Table 3 shows just how different these variances are, for the different priors and also for the MLE. Indeed, the variances virtually double with each step going from the MLE, to the Jeffreys prior, to $s = 3$, to $s = 2$, to $s = 1$. It will shortly be shown that the $s = 1$ reference prior yields procedures with good frequentist properties, so the variances for other priors and especially for the MLE estimates are likely to be severe underestimates of variability. This and high sensitivity of the analysis to the choice of the method and/or prior were also shown by Mitchell (1963). Of course, this sensitivity is not surprising, since each data set has only four observations and there are four unknown parameters.

When different conditional methods yield quite different answers, it is common to compare them through consideration of some frequentist criterion. Here we consider, for the Patterson model and design (but random data), the frequentist coverage of the interval from 0 to the 0.05 quantile (and from 0 to the 0.95 quantile) for ρ, the quantiles being those of the posterior distributions for the reference priors. (The computations were not done for the Jeffreys prior, because of difficulties caused by the extra integration needed in working with the Jeffreys prior; from Tables 2 and 3, one would expect the frequentist coverage probabilities of the 0.05 (0.95) quantile of the Jeffreys prior to be higher (lower) than those of the $s = 3$ reference prior.) Also included in the comparison were three priors from Mitchell (1963) that yield proper marginal posterior densities for ρ. These priors, denoted by g_1, g_2 and g_3, are $(1-\rho)/\sigma$, $\rho(1-\rho)^2/\sigma$ and $\rho^2(1-\rho)/\sigma$, respectively.

In this example, since there are 4 parameters in the model and only four observations, asymptotic classical confidence intervals are unusable (see, also , Mitchell, 1963). And because of the bounded parameter space $(0,1)$ of ρ and the presence of (α, β, σ) as nuisance parameters, there are no natural likelihood confidence sets for ρ. Hence no non-Bayesian confidence sets are studied here.

It is shown in Appendix C that frequentist coverage probabilities for intervals based on posterior quantiles using the reference priors or Jeffreys prior or Mitchell's priors depend only on $(\rho, |\beta|/\sigma)$. Table 4 presents these coverage probabilities for a variety of values of $(\rho, |\beta|/\sigma)$. Since ρ is a parameter in the bound set $(0,1)$, one cannot expect coverage probabilities and conditional probabilities to correspond for all parameter values; thus the coverage probabilities should be used merely to indicate which priors seem to yield consistently pessimistic or optimistic answers. (Of course , being consistently optimistic is probably worse.) In this respect, the reference priors for $s = 3$ and $s = 2$ and Mitchell's g_2 and g_3 seem inferior since they tend to yield too large (too small) coverage for the 0.05(0.95) tails. The clear winner among the reference priors appears to be the $s = 1$ reference prior, which yields coverages as large (small) as 0.12 (0.74) only rarely, and quite often yields coverages near the conditional 0.05 (0.95). Of considerable interest is that the $s = 1$ reference prior (which is

$(\rho, \lvert\beta\rvert/\sigma)$	$s=1$	$s=2$	$s=3$	g_1	g_2	g_3
(0.1, 0.1)	.05(1.00)	.10(1.00)	.13(1.00)	.05(1.00)	.78(1.00)	.997(1.00)
(0.1, 1)	.07(1.00)	.12(1.00)	.17(1.00)	.06(1.00)	.72(1.00)	.995(1.00)
(0.1, 10)	.12(.99)	.18(.96)	.22(.92)	.10(.99)	.28(.996)	.85(.998)
(0.1, 50)	.06(.94)	.15(.86)	.20(.79)	.06(.95)	.10(.97)	.27(.99)
(0.1, 100)	.06(.95)	.15(.81)	.19(.76)	.06(.95)	.08(.96)	.14(.99)
(0.1, 250)	.05(.95)	.14(.78)	.16(.71)	.05(.92)	.06(.96)	.02(.97)
(0.5, 0.1)	.002(.99)	.007(.98)	.016(1.00)	.001(.99)	.0008(1.00)	.015(1.00)
(0.5, 1)	.003(.99)	.013(.98)	.019(.96)	.003(.99)	.002(.99)	.02(1.00)
(0.5, 10)	.05(.94)	.14(.85)	.20(.80)	.05(.93)	.04(.94)	.16(.98)
(0.5, 50)	.05(.94)	.14(.84)	.20(.79)	.05(.94)	.05(.94)	.08(.96)
(0.5, 100)	.05(.95)	.14(.86)	.20(.80)	.05(.95)	.05(.95)	.07(.96)
(0.5, 250)	.05(.95)	.14(.86)	.19(.79)	.05(.95)	.05(.95)	.06(.95)
(0.9, 0.1)	0.(.79)	0.(.73)	0.(.67)	0.(.73)	0.(.02)	0.(.99)
(0.9, 1)	.0002(.78)	0.(.71)	0.0002(.69)	0.(.73)	0.(.02)	0.(.99)
(0.9, 10)	.0002(.74)	.001(.71)	.003(.67)	0.(.73)	0.(.1)	.0004(.97)
(0.9, 50)	.007(.84)	.02(.80)	.03(.75)	.004(.84)	.0006(.58)	.006(.92)
(0.9, 100)	.01(.90)	.06(.83)	.11(.78)	.01(.88)	.004(.72)	.02(.92)
(0.9, 250)	.04(.93)	.13(.85)	.19(.78)	.04(.92)	.02(.85)	.06(.93)

Table 4. Frequentist coverage probabilities of 0.05(0.95) posterior quantiles.

based on treating each parameter as a separate "grouping") is the reference prior recommended in Berger and Bernardo (1989a). Note that the coverages using g_1 are quite close to those using the reference prior for $s = 1$. In retrospect this is not surprising for the Patterson model because, from (26), it can be seen that g_1 and the $s = 1$ reference prior are proportional up to a smooth function that varies only between $\sqrt{1/2}$ and $\sqrt{1/3}$. The similarity here is something of a coincidence, however, as is indicated by (12) for the general case. Recall that one of our purposes was to test the recommended reference prior algorithm of Berger and Bernardo, and hence it appears to have been quite successful. One does not typically have the luxury of theoretically comparing a variety of noninformative priors; that the reference prior algorithm seems to automatically yield a trustworthy noninformative prior is thus encouraging.

Table 4 was done by simulation, generating 5,000 $\{y_i\}_{i=0}^{3}$ for each $(\rho, |\beta|/\sigma)$, calculating the indicated posterior quantiles using the reference priors for each set of generated data, and determining the proportion of quantiles that exceeded ρ. Calculations were done on a VAX11/780 computer and the standard error of each entry can be estimated by $\sqrt{p(1-p)/(5,000)}$ where p is the entry. (The entries are estimated to be accurate to within one unit in the last given digit.)

5 Conclusions

The first point to be made is that standard noninformative prior analyses, namely use of the Jeffreys or reference priors, do yield proper Bayesian inferences for the exponential regression problem. A second observation is that , as observed by Mitchell (1963), standard likelihood analysis here can lead to severe underestimation of error when there is little data. A third point concerns the possibility of using the grouped reference prior theory for robustness studies. By determining all grouped reference priors, and studying the differences among them, one has some feeling as to sensitivity of the conclusions to choice of noninformative prior. Finally, the numerical study provides some evidence that the reference prior based on treating each parameter as a separate group (i.e., $\sigma^{-1}\rho^{x-1}p(\rho^a)$) is the best reference prior; we, hence, recommend use of this prior for inferences concerning ρ in the exponential regression model.

References

[1] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

[2] Berger, J. & Bernardo, J. (1989a). Ordered group reference priors with applications to multinomial and variance components problems. Technical report **89-31c**, Purdue University, Dept. of Statistics.

[3] Berger, J. & Bernardo, J. (1989b). Estimating a product of means : Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200-207.

[4] Berger, J. & Bernardo, J. (1989c). Reference priors in a variance components problem. Technical report **89-32c**, Purdue University, Dept. of Statistics.

[5] Bernardo, J. (1977). Inferences about the ratio of normal means: a Bayesian approach to the Fieller-Creasy problem. In *Recent Development in Statistics (J.R. Barra et al., eds)*, 345-349. North-Holland, Amsterdam.

[6] Bernardo, J. (1979). Reference posterior distributions for Bayesian inference (with Discussion). *J.R. Statist. Soc.* **B 41**,113- 147.

[7] Consonni, G. & Veronese, P. (1989). Some remarks on the use of improper priors for the analysis of exponential regression models. *Biometrika* **76**, 101-106.

[8] Cox, D. & Hinkley, D. (1974). *Theoretical Statistics.* Chapman and Hall, London.

[9] Efron, B. (1986). Why isn't everyone a Bayesian?. *The American Statistician* **40**, 1-11.

[10] Jeffreys, H. (1961). *Theory of Probability (3rd ed.).* Oxford Univ. Press, London.

[11] Mitchell, A. (1963). The use of prior probabilities in estimating a parameter of the exponential regression equation. Ph.D thesis, Univ. of Cambridge.

[12] Mitchell, A. (1967). Discussion of paper by I.J. Good. *J.R. Statist. Soc.* **B 29**, 423-424.

[13] Patterson, H. (1960). A further note on a simple method for fitting an exponential curve. *Biometrika* **47**, 177-180.

[14] Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics* (R. Zieliński, ed.) 485-514. PWN-Polish Scientific Publishers, Warsaw.

Appendix A

First, we briefly sketch the algorithm from Berger and Bernardo (1989a) for computing ordered group reference priors. Let $\theta = (\theta_{(1)}, \theta_{(2)}, \cdots, \theta_{(m)})$ be the m groups of unknown parameters, where each $\theta_{(i)} = (\theta_{i_1}, \theta_{i_2}, \cdots, \theta_{i_{n_i}})$ has size n_i. Define

$$\theta_{[i]} = (\theta_{(1)}, \theta_{(2)}, \cdots, \theta_{(i)}), \text{ and } \theta_{[-i]} = (\theta_{(i+1)}, \theta_{(i+2)}, \cdots, \theta_{(m)}).$$

Let $\Sigma(\theta) = I^{-1}(\theta)$, where $I(\theta)$ is the Fisher information matrix, and define the decomposition of $\Sigma(\theta)$ corresponding to the $\theta_{(i)}$ as

$$\Sigma(\theta) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ & & \cdots & \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{pmatrix}.$$

Define

$$\Sigma_i = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1i} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2i} \\ & & \cdots & \\ \Sigma_{i1} & \Sigma_{i2} & \cdots & \Sigma_{ii} \end{pmatrix}, H_i = \Sigma_i^{-1} = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1i} \\ H_{21} & H_{22} & \cdots & H_{2i} \\ & & \cdots & \\ H_{i1} & H_{i2} & \cdots & H_{ii} \end{pmatrix},$$

and $h_i = \det(H_{ii})$.

Suppose $\Theta^1 \subset \Theta^2 \subset \cdots$ are a sequence of compact subsets of Θ such that $\cup_{i=1}^{\infty} \Theta^i = \Theta$ where Θ is the domain for θ. Define

$$\Theta^l_{\theta_{[i]}} = \{\theta_{(i+1)} : (\theta_{[i]}, \theta_{(i+1)}, \theta^*_{[-(i+1)]}) \in \Theta^l \; for \; some \; \theta^*_{[-(i+1)]}\}.$$

Also define

$$I_{\Omega}(x)=\begin{cases}1 & \text{if } y\in\Omega\\ 0 & \text{otherwise}\end{cases},$$

and $I^l(x|\theta_{[i]})=I_{\Theta^l_{\theta_{[i]}}}(x)$. The algorithm starts by defining

$$\pi^l_m(\theta_{[-(m-1)]}|\theta_{[m-1]})=\frac{|h_m(\theta)|^{\frac{1}{2}}I^l(\theta_{(m)}|\theta_{[m-1]})}{\int_{\Theta^l_{\theta_{[m-1]}}}|h_m(\theta)|^{\frac{1}{2}}I^l(\theta_{(m)}|\theta_{[m-1]})d\theta_{(m)}}.$$

It proceeds by iterating on the groups defining, for $i=m-1,m-2,\cdots,1$, π^l_i in terms of the previously defined π^l_j, $j=i+1,\cdots,m$, as

$$\pi^l_i(\theta_{[-(i-1)]}|\theta_{[i-1]})=\frac{\pi^l_{i+1}(\theta_{[-i]}|\theta_{[i]})\exp\{\frac{1}{2}E^l_\theta[\log|h_i(\theta)||\theta_{[i]}]\}I^l(\theta_{(i)}|\theta_{[i-1]})}{\int_{\Theta^l_{\theta_{[i-1]}}}\exp\{\frac{1}{2}E^l_\theta[\log|h_i(\theta)||\theta_{[i]}]\}I^l(\theta_{(i)}|\theta_{[i-1]})d\theta_{(i)}},$$

where

$$E^l_\theta[g(\theta)|\theta_{[i]}]=\int_{\{\theta_{[-i]}:(\theta_{[i]},\theta_{[-i]})\in\Theta^l\}}g(\theta)\pi^l_{i+1}(\theta_{[-i]}|\theta_{[i]})d\theta_{[-i]},$$

and one interprets $\theta_{[-0]}\hat{=}\theta$ and $\theta_{[0]}$ as vacuous, and writes $\pi^l(\theta)=\pi^l_1(\theta_{[-0]}|\theta_{[0]})$. The reference prior is then defined as

$$\pi(\theta)=\lim_{l\to\infty}\frac{\pi^l(\theta)}{\pi^l(\theta^*)},$$

provided it exists, where θ^* is some point in Θ^1.

To compute the reference priors in our problem, we first define

$$\Sigma(\rho,\alpha,\beta,\sigma)=H^{-1}(\rho,\alpha,\beta,\sigma),$$

where $H(\rho,\alpha,\beta,\sigma)$ is the Fisher information matrix (see (3)).

The determinant of H is

$$\det[H(\rho,\alpha,\beta,\sigma)]=\frac{2km}{\sigma^2}\det[H_1(\rho,\beta)]=\frac{2km^4}{\sigma^8}\beta^2\rho^{2(2x-1)}H_2(\rho), \tag{30}$$

where

$$
\begin{aligned}
H_2 &= \det\begin{pmatrix}
\sum_{i=0}^{k-1}(x+x_ia)^2\rho^{2ax_i} & \sum_{i=0}^{k-1}(x+x_ia)\rho^{ax_i} & \sum_{i=0}^{k-1}(x+x_ia)\rho^{2ax_i} \\
\sum_{i=0}^{k-1}(x+x_ia)\rho^{ax_i} & k & \sum_{i=0}^{k-1}\rho^{ax_i} \\
\sum_{i=0}^{k-1}(x+x_ia)\rho^{2ax_i} & \sum_{i=0}^{k-1}\rho^{ax_i} & \sum_{i=0}^{k-1}\rho^{2ax_i}
\end{pmatrix} \\
&= ka^2[(\sum_{i=0}^{k-1}\rho^{2ax_i} - \frac{1}{k}(\sum_{i=0}^{k-1}\rho^{ax_i})^2)(\sum_{i=0}^{k-1}x_i^2\rho^{2ax_i} - \frac{1}{k}(\sum_{i=0}^{k-1}x_i\rho^{ax_i})^2) \\
&\quad -(\sum_{i=0}^{k-1}x_i\rho^{2ax_i} - \frac{1}{k}\sum_{i=0}^{k-1}\rho^{ax_i}\sum_{i=0}^{k-1}x_i\rho^{ax_i})^2] = ka^2p_1^2(\rho^a).
\end{aligned} \tag{31}
$$

Defining

$$\Sigma = (\sigma_{ij}^*)_{4\times 4},$$

then

$$\det[H(\rho,\alpha,\beta,\sigma)]\sigma_{11}^* = \frac{m^3}{\sigma^6}\det\begin{pmatrix} k & \rho^x\sum_{i=0}^{k-1}\rho^{x_ia} & 0 \\ \rho^x\sum_{i=0}^{k-1}\rho^{x_ia} & \rho^{2x}\sum_{i=0}^{k-1}\rho^{2x_ia} & 0 \\ 0 & 0 & 2k \end{pmatrix} = \frac{2k^2m^3}{\sigma^6}\rho^{2x}p_2^2(\rho^a).$$

Hence

$$\sigma_{11}^* = \frac{\sigma^2}{a^2m\beta^2\rho^{2(x-1)}p^2(\rho^a)}.$$

Because we are only interested in inference for ρ, following the algorithm described in the beginning of the section, we compute the reference priors for ρ with respect to the ordered groups (1) $\{\rho,(\alpha,\beta,\sigma)\}$, (2) $\{\rho,\alpha,(\beta,\sigma)\}$ and (3) $\{\rho,\alpha,\beta,\sigma\}$. Other groups with permutations of α,β,σ are similar to one of the above three.

We will define the ranges for α,β,σ,ρ to be $\alpha\in[a_l,b_l]$, $\beta\in[c_l,d_l]$, $\sigma\in[e_l,f_l]$, $\rho\in[s_l,t_l]$, find the reference priors for these ranges, and then let $a_l,c_l\to-\infty, b_l,d_l,f_l\to+\infty, e_l,s_l\to 0$ and $t_l\to 1$ as $l\to\infty$. All K's used below represent constants.

(1) $\{\rho,(\alpha,\beta,\sigma)\}$:

$$h_1 = \sigma_{11}^{*\,-1}, h_2 = \frac{2k^2m^3\rho^{2x}}{\sigma^6}(\sum_{i=0}^{k-1}\rho^{2x_ia} - \frac{1}{k}(\sum_{i=0}^{k-1}\rho^{x_ia})^2) = \frac{2k^2m^3\rho^{2x}}{\sigma^6}p_2^2(\rho^a).$$

For $j = 2$:

$$\begin{aligned}
\pi_2^l((\alpha,\beta,\sigma)|\rho) &= \frac{|h_2|^{\frac{1}{2}} I_{(a_l,b_l)}(\alpha) I_{(c_l,d_l)}(\beta) I_{(e_l,f_l)}}{\int_{e_l}^{f_l}\int_{c_l}^{d_l}\int_{a_l}^{b_l} |h_2|^{\frac{1}{2}} d\alpha d\beta d\sigma} \\
&= \frac{\frac{1}{\sigma^3} I_{(a_l,b_l)}(\alpha) I_{(c_l,d_l)}(\beta) I_{(e_l,f_l)}(\sigma)}{(b_l - a_l)(d_l - c_l)\int_{e_l}^{f_l} \frac{1}{\sigma^3} d\sigma} \\
&= \frac{1}{\sigma^3} K_2(a_l,b_l,c_l,d_l,e_l,f_l) I_{(a_l,b_l)\times(c_l,d_l)\times(e_l,f_l)}(\alpha,\beta,\sigma).
\end{aligned}$$

Iteration for $j = 1$:

$$\begin{aligned}
E_\theta^l[\log|h_1(\alpha,\beta,\sigma,\rho)||\rho] &= \frac{\int_{c_l}^{d_l}\int_{e_l}^{f_l} \frac{1}{\sigma^3} \log({\sigma_{11}^*}^{-1}) d\sigma d\beta}{(d_l - c_l)(e_l^{-2} - f_l^{-2})} \\
&= \log(\rho^{2(x-1)} p^2(\rho^a)) + K_1(c_l,d_l,e_l,f_l).
\end{aligned}$$

Thus

$$\begin{aligned}
\pi^l(\alpha,\beta,\sigma,\rho) &= \pi_1^l(\alpha,\beta,\sigma,\rho) \\
&= \frac{\pi_2^l \exp\{\frac{1}{2}E_\theta^l[\log|h_1||\rho\} I_{(s_l,t_l)}(\rho)]}{\int_{s_l}^{t_l} \exp[\frac{1}{2}E_\theta^l\{\log|h_1||\rho]\} d\rho} I_{(a_l,b_l)\times(c_l,d_l)\times(e_l,f_l)\times(s_l,t_l)}(\alpha,\beta,\sigma,\rho) \\
&= \frac{1}{\sigma^3}\rho^{x-1} p(\rho^a) K(a_l,b_l,c_l,d_l,e_l,f_l,s_l,t_l) I_{(a_l,b_l)\times(c_l,d_l)\times(e_l,f_l)\times(s_l,t_l)}(\alpha,\beta,\sigma,\rho).
\end{aligned}$$

Therefore, the reference prior in this group is

$$\begin{aligned}
\pi(\{\rho,(\alpha,\beta,\sigma)\}) &\propto \lim_{l\to\infty} \frac{\pi^l(\alpha,\beta,\sigma,\rho)}{\pi^l(0,0,1,\frac{1}{2})} \\
&\propto \frac{1}{\sigma^3}\rho^{x-1} p(\rho^a).
\end{aligned}$$

(2) $\{\rho,\alpha,(\beta,\sigma)\}$:

$$h_1 = {\sigma_{11}^*}^{-1},\ h_2 = \frac{\sigma_{11}^*}{\sigma_{11}^*\sigma_{22}^* - {\sigma_{12}^*}^2},\ h_3 = \frac{2km^2\rho^{2x}}{\sigma^4}\sum_{i=0}^{k-1}\rho^{2x_i a}.$$

Note that $h_2 = \sigma_1(\rho)/\sigma^2$, where

$$\sigma_1(\rho) = mp_2^2(\rho^a)/[k\sum_{i=0}^{k-1}\rho^{2x_i a}].$$

Now, for $j = 3$:

$$\begin{aligned}
\pi_3^l((\beta,\sigma)|\rho,\alpha) &= \frac{|h_3|^{\frac{1}{2}} I_{(c_l,d_l)}(\beta) I_{(e_l,f_l)}(\sigma)}{\int_{e_l}^{f_l}\int_{c_l}^{d_l} |h_3|^{\frac{1}{2}} d\beta d\sigma} \\
&= \frac{\frac{1}{\sigma^2} I_{(c_l,d_l)}(\beta) I_{(e_l,f_l)}(\sigma)}{(d_l - c_l)(e_l^{-1} - f_l^{-1})}.
\end{aligned}$$

Iteration for $j = 2$:

$$E_\theta^l[\log|h_2||\alpha,\rho] = \frac{\int_{e_l}^{f_l} \frac{1}{\sigma^2}\log|h_2|d\sigma}{e_l^{-1} - f_l^{-1}},$$

so

$$\pi_2^l(\alpha,(\beta,\sigma)|\rho) \propto \frac{\frac{1}{\sigma^2} I_{(a_l,b_l)}(\alpha) I_{(c_l,d_l)}(\beta) I_{(e_l,f_l)}}{(b_l - a_l)(d_l - c_l)(e_l^{-1} - f_l^{-1})}.$$

Iteration for $j = 1$:

$$\begin{aligned} E_\theta^l[\log|h_1||\rho] &= \frac{\int_{e_l}^{f_l}\int_{c_l}^{d_l} \frac{1}{\sigma^2}\log|h_1|d\beta d\sigma}{(d_l - c_l)(e_l^{-1} - f_l^{-1})} \\ &= \log(\rho^{2(x-1)} p^2(\rho^a)) + K_1'(c_l, d_l, e_l, f_l). \end{aligned}$$

Thus

$$\pi_1^l(\alpha,\beta,\sigma,\rho) = \frac{1}{\sigma^2}\rho^{x-1} p(\rho^a) K'(a_l,b_l,c_l,d_l,e_l,f_l,s_l,t_l) I_{(a_l,b_l)\times(c_l,d_l)\times(e_l,f_l)\times(s_l,t_l)}(\alpha,\beta,\sigma,\rho).$$

Therefore, the reference prior is

$$\begin{aligned} \pi(\{\rho,\alpha,(\beta,\sigma)\}) &\propto \lim_{l\to\infty} \frac{\pi^l(\alpha,\beta,\sigma,\rho)}{\pi^l(0,0,1,\frac{1}{2})} \propto \lim_{l\to\infty} \frac{\pi_1^l(\alpha,\beta,\sigma,\rho)}{\pi_1^l(0,0,1,\frac{1}{2})} \\ &\propto \frac{1}{\sigma^2}\rho^{x-1} p(\rho^a). \end{aligned}$$

(3) $\{\rho,\alpha,\beta,\sigma\}$:

$$h_1 = {\sigma_{11}^*}^{-1}, h_2 = \frac{\sigma_1(\rho)}{\sigma^2}, h_3 = \frac{m\rho^{2x}}{\sigma^2}\sum_{i=0}^{k-1}\rho^{2x_i a}, h_4 = \frac{2km}{\sigma^2}.$$

For $j = 4$:

$$\pi_4^l(\sigma|\rho,\alpha,\beta) = \frac{|h_4|^{\frac{1}{2}} I_{(e_l,f_l)}(\sigma)}{\int_{e_l}^{f_l}|h_4|^{\frac{1}{2}}d\sigma} = \frac{\frac{1}{\sigma}I_{(e_l,f_l)}(\sigma)}{\log(e_l^{-1}f_l)}.$$

Iteration for $j = 3$:

$$E_\theta^l[\log|h_3||\rho,\alpha,\beta] = \frac{\int_{e_l}^{f_l}\frac{1}{\sigma}\log|h_3|d\sigma}{\log(e_l^{-1}f_l)} = \log(\rho^{2x}\sum_{i=0}^{k-1}\rho^{2x_i a}) + K_3''(e_l, f_l),$$

so

$$\pi_3^l(\beta,\sigma|\rho,\alpha) \propto \frac{\frac{1}{\sigma}I_{(c_l,d_l)}(\beta) I_{(e_l,f_l)}(\sigma)}{(d_l - c_l)\log(e_l^{-1}f_l)},$$

since $E_\theta^l[\log|h_3||\rho,\alpha,\beta]$ does not depend on β. Iteration for $j=2$: since h_2 does not depend on α, it follows that

$$\pi_2^l(\alpha,\beta,\sigma|\rho)\propto\frac{\frac{1}{\sigma}I_{(a_l,b_l)}(\alpha)I_{(c_l,d_l)}(\beta)I_{(e_l,f_l)}(\sigma)}{(b_l-a_l)(d_l-c_l)\log(e_l^{-1}f_l)}.$$

Iteration for $j=1$: as in case (2), we have

$$\begin{aligned}E_\theta^l[\log|h_1||\rho] &= \frac{\int_{e_l}^{f_l}\int_{c_l}^{d_l}\frac{1}{\sigma^2}\log|h_1|d\beta d\sigma}{(d_l-c_l)(e_l^{-1}-f_l^{-1})}\\ &= \log(\rho^{2(x-1)}p^2(\rho^a))+K_1''(c_l,d_l,e_l,f_l).\end{aligned}$$

Therefore, as in case (2),

$$\pi(\{\rho,\alpha,\beta,\sigma\})\propto\frac{1}{\sigma}\rho^{x-1}p(\rho^a).$$

Appendix B

In this Appendix, Taylor expansion will be used to approximate $p_1^2(\rho)$ and $p_2^2(\rho)$ near the points 0 and 1. Let

$$\begin{aligned}f_1(\rho) &= \sum_{i=0}^{k-1}\rho^{2x_i}-\frac{1}{k}(\sum_{i=0}^{k-1}\rho^i)^2,\\ f_2(\rho) &= \sum_{i=0}^{k-1}x_i\rho^{2x_i}-\frac{1}{k}\sum_{i=0}^{k-1}\rho^i\sum_{i=0}^{k-1}x_i\rho^{x_i},\\ f_3(\rho) &= \sum_{i=0}^{k-1}x_i^2\rho^{2x_i}-\frac{1}{k}(\sum_{i=0}^{k-1}x_i\rho^{x_i})^2.\end{aligned}\tag{32}$$

(a) For $\rho\sim 0$: Since

$$\begin{aligned}f_1(\rho) &= p_2^2(\rho)=\rho^{2x_0}+\rho^{2x_1}+o(\rho^{2x_1})-\frac{1}{k}(\rho^{x_0}+\rho^{x_1}+\rho^{x_2}+o(\rho^3))^2\\ &= (1-\frac{1}{k})\rho^{2x_0}+(1-\frac{1}{k})\rho^{2x_1}-\frac{2}{k}\rho^{x_0+x_1}-\frac{2}{k}\rho^{x_0+x_2}+o(\rho^{\min\{x_0+x_2,2x_1\}}),\end{aligned}\tag{33}$$

$$\begin{aligned}f_2(\rho) &= x_0\rho^{2x_0}+x_1\rho^{2x_1}-\frac{1}{k}(\rho^{x_0}+\rho^{x_1}+\rho^{x_2}+o(\rho^{x_2})(x_0\rho^{x_0}+x_1\rho^{x_1}+x_2\rho^{x_2}+o(\rho^{x_2}))\\ &= x_0(1-\frac{1}{k})\rho^{2x_0}+x_1(1-\frac{1}{k})\rho^{2x_1}-\frac{x_0+x_1}{k}\rho^{x_0+x_1}-\frac{x_0+x_2}{k}\rho^{x_0+x_2}+o(\rho^{\min\{x_0+x_2,2x_1\}}),\end{aligned}$$

and

$$
\begin{aligned}
f_3(\rho) &= x_0^2\rho^{2x_0} + x_1^2\rho^{2x_1} + o(\rho^{2x_1}) - \frac{1}{k}(x_0\rho^{x_0} + x_1\rho^{x_1} + x_2\rho^{x_2} + o(\rho^{x_2}))^2 \\
&= x_0^2(1-\frac{1}{k})\rho^{2x_0} + x_1^2(1-\frac{1}{k})\rho^{2x_1} - \frac{2x_0x_1}{k}\rho^{x_0+x_1} - \frac{2x_0x_2}{k}\rho^{x_0+x_2} + o(\rho^{\min\{x_0+x_2,2x_1\}}),
\end{aligned}
$$

it follows that

$$
\begin{aligned}
p_1^2(\rho) &= f_1(\rho)f_3(\rho) - f_2^2(\rho) \\
&= (1-\frac{2}{k})(x_1-x_0)^2\rho^{2x_0+2x_1} + o(\rho^{2x_0+2x_1}). \qquad (34)
\end{aligned}
$$

(b) For $\rho \sim 1$: let $r = 1-\rho$, and observe that

$$
\begin{aligned}
p_2^2(\rho) &= \sum_{i=0}^{k-1}(1-r)^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1}(1-r)^{x_i})^2 \\
&= \sum_{i=0}^{k-1}(1-2x_ir + x_i(2x_i-1)r^2) \\
&\quad -\frac{1}{k}\sum_{i=0}^{k-1}(1-x_ir+\frac{x_i(x_i-1)}{2}r^2)^2 + o(r^2) \\
&= [\sum_{i=0}^{k-1}x_i^2 - \frac{1}{k}(\sum_{i=0}^{k-1}x_i)^2]r^2 + o(r^2) = qr^2 + o(r^2). \qquad (35)
\end{aligned}
$$

To determine the behavior of $p_1^2(\rho)$ at $\rho \sim 1$, consider a Taylor expansion of $p_1^2(\rho)$. Since $p_1^2(\rho)$ is a polynomial, it can be expanded as

$$
p_1^2(\rho) = p_1^2(1) + \frac{d}{d\rho}p_1^2(1)r + \frac{1}{2!}\frac{d^2}{d\rho^2}p_1^2(1)r^2 + \frac{1}{3!}\frac{d^3}{d\rho^3}p_1^2(1)r^3 + \frac{1}{4!}\frac{d^4}{d\rho^4}p_1^2(1)r^4 + o(r^4).
$$

Calculation shows that $p_1^2(1)$ and the first three derivatives are all zero when $\rho = 1$, and $\frac{d^4}{d\rho^4}p_1^2(1) = 6Q$ (see (7)). Thus

$$
p_1^2(\rho) = \frac{1}{4}Qr^4 + o(r^4). \qquad (36)
$$

Appendix C

Here we show that the frequentist coverage probabilities for intervals based on the posterior quantiles with respect to the reference priors or the Jeffreys prior depend only on $(\rho, |\beta|/\sigma)$.

Definition: *A function $f(\mathbf{x})$ is called a homogeneous function if for any real constant c, $f(c\mathbf{x}) = f(\mathbf{x})$.*

Lemma 1 *Suppose a function g of $x_1, x_2, \cdots, x_{k-1}$, which is denoted by $g(\{x_i\}_{i=1}^{k-1})$, is a homogeneous function. Then the probability $P(g)=P_{(\rho,\alpha,\beta,\sigma)}(\rho \le g(\{\overline{Y}_{i.} - \overline{Y}\}_{i=1}^{k-1}))$ depends only on $(\rho, |\beta|/\sigma)$ for the exponential regression model (1).*

Proof: The density function of $\mathbf{Y} = \{Y_{ij}\}$ for $0 \le i \le k-1$ and $1 \le j \le m$ is

$$\begin{aligned} f(\mathbf{y}|\alpha,\beta,\sigma,\rho) &= C_{k,m}\sigma^{-km}(s^2)^{\frac{km-k-2}{2}} \\ &\times \exp\{-\frac{1}{2\sigma^2}[s^2 + m\sum_{i=0}^{k-1}(\overline{y}_{i.} - \overline{y} - \beta(\rho^{x+ax_i} - \overline{\rho}))^2 + km(\overline{y} - \alpha - \beta\overline{\rho})^2]\}, \quad (37) \end{aligned}$$

where $\overline{\rho} = \frac{\rho^x}{k}\sum_{i=0}^{k-1}\rho^{ax_i}$ and $C_{k,m}$ is a constant.

To rewrite $P(g)$, first make the transformation $u_0 = \overline{y}/\sigma$, $u_i = \text{sign}(\beta)(\overline{y}_{i.} - \overline{y})/\sigma$ for $1 \le i \le k-1$ and $u_k = s^2/\sigma^2$. The Jacobian is $k\sigma^{k+2}$. Now $g(\{\overline{y}_{i.} - \overline{y}\}_{i=0}^{k-2}) = g(\text{sign}(\beta)\sigma\{u_i\}_{i=1}^{k-1}) = g(\{u_i\}_{i=1}^{k-1})$ by homogeneity. Therefore

$$\begin{aligned} P(g) &= kC_{k,m}\sigma^{-km}\int_{\rho \le g(\{u_i\}_{i=1}^{k-1})}(u_k\sigma)^{\frac{km-k-2}{2}}\sigma^{k+2} \\ &\times \exp\{-\frac{1}{2}[s_u^2 + m\sum_{i=0}^{k-1}(u_i - \frac{|\beta|}{\sigma}(\rho^{x+ax_i} - \overline{\rho}))^2 + km(u_0 - \alpha - \frac{|\beta|}{\sigma}\overline{\rho})^2]\}du_0\cdots du_k \\ &= C_{k,m}^1\int_{\rho \le g(\{u_i\}_{i=1}^{k-1})}(u_k)^{\frac{km-k-2}{2}}\exp\{-\frac{1}{2}[u_k + m\sum_{i=0}^{k-1}(u_i - \frac{|\beta|}{\sigma}(\rho^{x+ax_i} - \overline{\rho}))^2]\}du_1\cdots du_k (38) \end{aligned}$$

where $C_{k,m}^1$ is a constant. It is then obvious that $P(g)$ depends only on $(\rho, |\beta|/\sigma)$. □

Lemma 2 *Suppose that $F(\rho)$ is a marginal posterior distribution for ρ using the reference priors or the Jeffreys prior or Mitchell's priors which are given in Section 4. Then the γth posterior quantile, ρ_γ, is a function $g(\{\overline{y}_{i.} - \overline{y}\}_{i=1}^{k-1})$, where g is a homogeneous function.*

Proof: (a) For the marginal posterior distribution of ρ, from (17), (18) and (19), it is easy to see that ρ_γ depends only on $\{\overline{y}_{i.} - \overline{y}\}_{i=0}^{k-1}$. The homogeneity is obvious since $h(\rho; s, a\mathbf{y}) = a^{km+s-3}h(\rho; s, \mathbf{y})$, and the same factor of a arises from the normalization constant $K(s)$; hence, the factors of a can be cancelled.

(b) For the Jeffreys prior or Mitchell's priors, the proof is similar. $\square$

Appendix D

This appendix gives the forms of the estimated asymptotic variances of the MLE using the observed Fisher information matrix and expected Fisher information matrix, defined as MLE(1) and MLE(2), respectively, in Table 3.

Let $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\rho, \alpha, \beta, \sigma)$ denote the parameters, and $f(\mathbf{y}|\theta)$ be the likelihood function. The asymptotic variances associated with MLE(1) and MLE(2), respectively, are given by the $(4,4)$ element of V, where (cf. Berger, 1985)

(1) for MLE(1): $V = [\hat{I}(\mathbf{y})]^{-1}$, where

$$\hat{I}_{uv}(\mathbf{y}) = -\sum_{i=0}^{k-1} \sum_{j=1}^{m} [\frac{\partial^2}{\partial\theta_u \partial\theta_v} \log f(y_{ij}|\theta)]_{\theta=\hat{\theta}},$$

for $1 \leq u, v \leq 4$, and $\hat{\theta}$ is the MLE of the parameters.

(2) for MLE(2): $V = [I(\hat{\theta})]^{-1}$, where, for $1 \leq u, v \leq 4$,

$$I_{uv}(\theta) = -m \sum_{i=0}^{k-1} E_\theta^{Y_{i1}} [\frac{\partial^2}{\partial\theta_u \partial\theta_v} \log f(Y_{i1}|\theta)].$$