Biometrika (1981), **68**, 1, *pp.* 35–43
Printed in Great Britain

A Bayesian approach to transformations to normality

BY L. R. PERICCHI

Programa de Postgrado en Recursos Hidricos, Universidad Simon Bolivar, Caracas, Venezuela

SUMMARY

The analysis of transformation of observations in the linear model with normal errors proposed by Box & Cox (1964) is considered. A different choice of noninformative unnormed prior is advocated, which is not outcome dependent. This new selection of prior leads to a formal identity between likelihood and Bayesian inference, both for the estimation of the best transformation to normality and for the presence of homoscedasticity and additivity under this transformation. Extension to a related problem is mentioned.

Some key words: Jeffreys's multiparameter prior; Outcome-dependent prior; Posterior model odds; Transformation to normality.

1. INTRODUCTION

The use of a parametric family of power transformations of data to obtain a simple linear normal model was studied by Box & Cox (1964), who developed Bayesian and likelihood arguments in parallel.

We shall be concerned here with two problems in the foundation of their analyses. First, the results obtained by the Bayesian and likelihood approaches could lead to contradictory inferences. Secondly, the improper 'noninformative' priors that Box & Cox introduced are 'outcome dependent', that is they depend on the outcome of the experiment, the observed vectors of responses. On the first point it is natural to enquire which inference should be followed in the case of opposed conclusions from their Bayesian and likelihood analyses. An example is given in this paper to show that such a case could well occur in practice. The second aspect has been criticized by several writers including Nelder (1964) and Fraser (1967, 1968). Box & Cox themselves repeatedly warned in their paper that there may be a better approach to the prior density than theirs, but to the best of our knowledge an alternative has not been proposed until now. About this Lindley (1972, p. 48) has stated that the Box & Cox analysis of transformations

> ... is puzzling because it uses a density over the parameter space, which depends on the data. This difficulty deserves more attention but appears to arise from the use of improper distribution.

Our chief claim is that this difficulty is not due to the use of improper distributions in general. It is the purpose of the present paper to introduce a different and sounder noninformative prior which does not depend on the observations. Furthermore, it leads to the same inferences as those based on maximum likelihood methods.

In § 2 the alternative prior density is motivated on various grounds, and the criteria derived for estimating the best power transformations that simultaneously approximates normality, homoscedasticity and additivity. In § 3 the analysis is extended to include the separate criteria of normality, homoscedasticity and additivity. An undesirable feature of the Box & Cox Bayesian statistic for checking additivity is pointed out; this feature is not shared by the statistic derived with the new prior. By means of a simulation following one of the authors' examples, it is shown how this feature is responsible for the discrepancy between Box & Cox's Bayesian and likelihood inference for additivity. In § 4 a related problem is briefly examined and a summary conclusion is given in § 5.

2. Derivation of the estimate for the power transformation

2·1. *Derivation of the likelihood estimate for* λ

Let $Y = (y_1, \ldots, y_n)$ be n independent observations. Box & Cox studied the following family of power transformations, for $y > 0$,

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & (\lambda \neq 0), \\ \log y & (\lambda = 0), \end{cases}$$

and for some λ to a sufficient approximation $Y^{(\lambda)} = X\theta + \varepsilon$, where $Y^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_n^{(\lambda)})$, ε is an $n \times 1$ vector of errors with distribution $N_n(0, \sigma^2 I)$, θ a $k \times 1$ vector of parameters and X an $n \times k$ matrix of fixed elements with rank k and first column of ones.

Let us suppose initially that we can achieve simultaneously normality, N, homoscedasticity, H, and additivity, A, with a particular power λ.

The likelihood function relative to the original untransformed observations Y can be written as

$$p_1(Y|\lambda, \theta, \sigma) = \frac{1}{(2\pi)^{\frac{1}{2}n}\sigma^n}\exp\left\{-\frac{S_\lambda + (\theta-\hat{\theta}_\lambda)'X'X(\theta-\hat{\theta}_\lambda)}{2\sigma^2}\right\}J_\lambda, \tag{2·1}$$

where

$$J_\lambda = \Pi_i y_i^{\lambda-1}, \quad \hat{\theta}_\lambda = (X'X)^{-1}X'Y^{(\lambda)}, \quad S_\lambda = (Y^{(\lambda)} - X\hat{\theta}_\lambda)'(Y^{(\lambda)} - X\hat{\theta}_\lambda).$$

For a fixed λ the maximized log likelihood, except for a constant common for every λ, is

$$L_{\max}^{(\lambda)} = -\tfrac{1}{2}n \log\{S_\lambda/(nJ_\lambda^{2/n})\}. \tag{2·2}$$

The estimate $\hat{\lambda}$ will be obtained by maximizing $L_{\max}^{(\lambda)}$.

2·2. *Posterior distribution of* λ

For the corresponding Bayesian analysis the prior distribution assumed by Box & Cox for $\lambda = 1$ is

$$p_0(\theta, \sigma, \lambda = 1) = p_0(\theta, \sigma | \lambda = 1)p_0(\lambda = 1) \propto p_0(\lambda = 1)/\sigma. \tag{2·3}$$

On the other hand, we are going to assume for $\lambda = 1$

$$\tilde{p}_0(\theta, \sigma, \lambda = 1) \propto I_n(\theta, \sigma)^{\frac{1}{2}} p_0(\lambda = 1) \propto p_0(\lambda = 1)/\sigma^{k+1}, \tag{2·4}$$

where $I_n(\theta, \sigma)$ is the Fisher information matrix.

The prior (2·4) is assigned according to Jeffreys's multiparameter rule, without the assumption of *a priori* independence between the parameters, an assumption that is made in (2·3).

To complete the choice of the prior for arbitrary λ, it has to be realized that the transformation will contract or enlarge the range of the transformed observations, a fact that could change the value of the priors.

To determine their prior Box & Cox argue as follows. Suppose that the power transformation is approximately linear over the range of observations, then

$$E(y_i^{(\lambda)}) \simeq a_\lambda + l_\lambda E(y_i), \tag{2·5}$$

where l_λ is some representative of the gradient $dy^{(\lambda)}/dy$.

Beginning with (2·3), a compatible assessment of the overall prior using (2·5) is

$$p_0(\theta, \sigma, \lambda) \propto p_0(\lambda)/(\sigma\,|\,l_\lambda|^k). \tag{2·6}$$

By the same reasoning, the alternative choice (2·4) will result in

$$\tilde{p}_0(\theta, \sigma, \lambda) \propto p_0(\lambda)/\sigma^{k+1}. \tag{2·7}$$

Therefore, if we use Box & Cox's approximation the shape of the alternative choice of prior $\tilde{p}_0(\theta, \sigma\,|\,\lambda)$ could be taken independent of λ, and independent of the outcome represented in (2·6) by l_λ. To specify (2·6) completely, a choice for l_λ has to be made. Box & Cox choose l_λ as the geometric mean of the Jacobian

$$l_\lambda = J_\lambda^{1/n} = (\Pi y_i^{\lambda-1})^{1/n}. \tag{2·8}$$

This choice for l_λ is qualified by Box & Cox as '. . . somewhat arbitrary'. Finally, from (2·8) and (2·6), Box & Cox's prior is

$$p_0(\theta, \sigma, \lambda) = p_0(\lambda)/(\sigma J_\lambda^{k/n}). \tag{2·9}$$

We call (2·9) an 'outcome-dependent' prior as opposed to a 'data-dependent' prior, because the latter more general term is also used to denote priors that depend merely on an ancillary statistic, for example n, and not necessarily on the actual outcome of the experiment.

The prior (2·9) combined with the likelihood (2·1) gives the posterior distribution

$$p_2(\theta, \sigma, \lambda\,|\,Y) \propto \sigma^{-n-1} \exp[-\{S_\lambda + (\theta-\hat{\theta}_\lambda)'X'X(\theta-\hat{\theta}_\lambda)\}/(2\sigma^2)]\,J_\lambda^{(n-k)/n} p_0(\lambda), \tag{2·10}$$

and for the alternative prior (2·7)

$$\tilde{p}_2(\theta, \sigma, \lambda\,|\,Y) \propto \sigma^{-n-k-1} \exp[-\{S_\lambda + (\theta-\hat{\theta}_\lambda)'X'X(\theta-\hat{\theta}_\lambda)\}/(2\sigma^2)]\,J_\lambda p_0(\lambda). \tag{2·11}$$

From now on we suppose $p_0(\lambda)$ uniform over Λ, the region of λ's under consideration. Integration of (2·10) with respect to θ and σ gives, except for a constant independent of λ and Y, if we take logs,

$$L_b(\lambda) = -\tfrac{1}{2}\nu_r \log\{S_\lambda/(\nu_r J_\lambda^{2/n})\}, \tag{2·12}$$

where $\nu_r = n-k$. Similarly, integration of (2·11) gives

$$\tilde{L}_b(\lambda) = -\tfrac{1}{2}n \log\{S_\lambda/(n J_\lambda^{2/n})\}. \tag{2·13}$$

From (2·2) and (2·13) we see that the likelihood and Bayesian statistics which have to be maximized for estimation of λ are identical when the alternative prior (2·7) is used. To proceed, let $\mathrm{HPD}_\alpha^\lambda$ be the highest posterior density interval for λ of level α for the prior (2·9) and let $\mathrm{HPD}_\alpha^{*\lambda}$ be similarly defined for the alternative (2·7). From the approximate normality of the posteriors densities for λ we have that

$$\mathrm{HPD}_\alpha^\lambda = \{\lambda\colon \log(S_\lambda/J_\lambda^{2/n}) - \log(S_{\hat{\lambda}}/J_{\hat{\lambda}}^{2/n}) < \chi_1^2(\alpha)/(n-k)\}, \tag{2·14}$$

and $\mathrm{HPD}_\alpha^{*\lambda}$ is the same as (2·14) except that $(n-k)$ is replaced by n. Therefore, $\mathrm{HPD}_\alpha^{*\lambda} \subset \mathrm{HPD}_\alpha^\lambda$. On the other hand, using the large-sample chi-squared distribution of the log likelihood ratio it follows from (2·2) that the approximate level α confidence interval for λ is formally equal to $\text{HPD}^{*\lambda}_{\alpha}$. Its difference from $\text{HPD}^{\lambda}_{\alpha}$ is appreciable when k is not negligible with respect to n.

2·3. *Different approaches to the alternative prior*

Besides the approximate compatibility argument given in the previous section, the alternative prior (2·7) can be motivated on other grounds. First, let us pretend that we have prior information for the parameters, in the form of a multinormal inverse gamma (Raiffa & Schlaifer, 1961, p. 343)

$$p_0(\theta\,|\,\sigma,\lambda) = \frac{|V_\lambda|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}\sigma^k}\exp\{-\tfrac{1}{2}\sigma^{-2}(\theta-\bar{\theta}_\lambda)'V_\lambda(\theta-\bar{\theta}_\lambda)\},\quad p_0(\sigma\,|\,\lambda) = \frac{K_\lambda}{\sigma^{\nu_\lambda+1}}\exp\left(-\frac{v_\lambda \bar{S}^2_\lambda}{2\sigma^2}\right). \tag{2·15}$$

We combine (2·15) with the likelihood (2·1) obtaining a joint posterior density for θ, σ and λ. If we next let $V_\lambda \to 0$ and $\nu_\lambda \to 0$ for all λ, approaching a noninformative situation, we get the posterior density (2·11), the same as starting with the alternative noninformative prior (2·7).

Secondly, we restrict the set of possible priors to those that produce large-sample posterior odds for different models, invariant with respect to changes in scale on the dependent as well as independent variables.

If we use results from Lindley (1961), the posterior odds of the model λ_1 over λ_2 are asymptotically

$$\frac{p_2(\lambda_1\,|\,Y)}{p_2(\lambda_2\,|\,Y)} = \frac{p_1(Y,\hat{\alpha}\,|\,\lambda_1)\,\xi_\alpha^{-\frac{1}{2}}\,p_0(\hat{\alpha}\,|\,\lambda_1)}{p_1(Y,\hat{\beta}\,|\,\lambda_2)\,\xi_\beta^{-\frac{1}{2}}\,p_0(\hat{\beta}\,|\,\lambda_2)}, \tag{2·16}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimators, for example $\hat{\alpha} = (\hat{\theta}_{\lambda_1}, \hat{\sigma}_{\lambda_1})$, and ξ_α and ξ_β are the information determinants for the estimation of α and β respectively from Y.

From the requirement of scale invariance on Y, it turns out, using (2·16), that the priors have to obey a homogeneous equation of order $-(k+1)$. Using Euler's theorem, this leads to a partial differential equation of Lagrange type with infinite solutions. If, furthermore, we require that the priors are constant with respect to changes in θ we get the necessary condition

$$\frac{\sigma\, dp_0(\sigma,\theta\,|\,\lambda)}{d\sigma} = -(k+1)\,p_0(\sigma,\theta\,|\,\lambda),$$

which has the unique solution $p_0(\sigma,\theta\,|\,\lambda) \propto 1/\sigma^{(k+1)}$, again the alternative prior (2·7). Moreover, (2·7) not only gives asymptotic scale invariance for changes in the scale of Y but also exact scale invariance, since it follows that the actual expression for the posterior odds is

$$\frac{\tilde{p}_2(\lambda_1\,|\,cY)}{\tilde{p}_2(\lambda_2\,|\,cY)} = (J_{\lambda_1}/J_{\lambda_2})\,(S_{\lambda_1}/S_{\lambda_2})^{\frac{1}{2}n}, \tag{2·17}$$

for all $c > 0$. Furthermore, this last expression is also invariant with respect to changes in scale of the dependent variable X.

Finally, on general grounds, Cox (1977) has asserted that when the number of adjustable parameters for different models is equal, as in the present case, there are some reasons to expect that, approximately, the product of the last two factors in (2·16) is unity. For the alternative prior (2·7) their product is exactly unity. On the other hand, for Box & Cox's prior (2·9) their product is

$$\left\{\frac{S_{\lambda_1}/J_{\lambda_1}}{S_{\lambda_2}/J_{\lambda_2}}\right\}^{\frac{1}{2}k}. \tag{2·18}$$

This expression is in turn related to the likelihood ratio. Call $A_r(\lambda_1, \lambda_2)$ the set of y's where the likelihood ratio is equal to r. For the prior (2·9) the exact posterior odds for model λ_1 versus λ_2 on the set $A_r(\lambda_1, \lambda_2)$ are equal to $r^{(n-k)/n}$, and the factor (2·18) is equal to $r^{-k/n}$. Again, if k/n is not negligible neither is (2·18) for particular values of λ_1 and λ_2. However, if the alternative prior (2·7) is assumed, the posterior odds are exactly equal to the likelihood ratio of the two models, as is seen from (2·17).

3. Further analysis of the transformation: Additivity and homogeneity of variance

3·1. *Constrained models*

We now look into the likelihood and Bayesian analysis of constraints, considered by Box & Cox, that allow us to perform the separated analysis of the different effects of normality, homoscedasticity and additivity. Considering a general model to which a constraint C could be applied or relaxed, it follows that

$$L_{\max}(\lambda \mid C) = L_{\max}(\lambda) + \{L_{\max}(\lambda \mid C) - L_{\max}(\lambda)\}, \tag{3·1}$$

and the second term on the right-hand side is a statistic to test the presence of the constraint.

For the corresponding Bayesian analysis of constrained models, the ratio of posterior densities is considered, namely

$$K_C(\lambda) = \frac{p_2(\lambda \mid C, Y)}{p_2(\lambda \mid Y)}. \tag{3·2}$$

We will be willing to accept the simpler model if high values of (3·2) are obtained in the region of λ's where the complex model has been shown adequate.

The interpretation of (3·2) should not be made in terms of the probability that the constraint C is true. It is not legitimate to use the expression for the conditional probability, as in formula (29) of Box & Cox. We will see in §§ 3·2 and 3·3 that the prior densities used, both in Box & Cox's set-up and in the present one for the simpler model, are not the same priors used for the complex model conditioned on C.

Therefore we will be comparing different models and so the interpretation of (3·2) will be made in terms of the posterior odds of the two models. The indifference value for $K_C(\lambda)$ is set at one. Equations (3·1) and (3·2) can be extended for a hierarchy of constraints C_1, C_2, etc.

3·2. *Criteria for additivity*

Let us now suppose at this point that with the transformation λ normality, N, and homoscedasticity, H, are achieved. A statistic should be derived to test if additivity, A, is reasonably attained in the region of λ's where the first two properties are true to a sufficient approximation.

Let $\theta = (\theta_1, \theta_2)$, where θ_2 is the interaction parameters with $\dim(\theta_1) = \nu_1$, $\dim(\theta_2) = \nu_2$, $\nu_1 + \nu_2 = k$ and let the residual sum of squares be $S^{(\lambda)}_{\nu_r}$ and $S^{(\lambda)}_{\nu_r+\nu_2}$ for the complex and simpler model, identified by their degrees of freedom, where $\nu_r = n-k$. It follows that, except for a constant,

$$\{L_{\max}(\lambda \mid H, N, A) - L_{\max}(\lambda \mid H, N)\} = -\tfrac{1}{2}n \log\left\{1+\frac{\nu_2}{\nu_r}F(\lambda)\right\}, \qquad (3\cdot3)$$

where $F(\lambda)$ is the usual F statistic.

For the Bayesian counterpart, Box & Cox set the following 'outcome-dependent' prior densities for the complex and simpler models respectively

$$p_0(\theta, \sigma, \lambda \mid N, H) \propto 1/(\sigma J_\lambda^{k/n}), \quad p_0(\theta_1, \sigma, \lambda \mid A, N, H) \propto 1/(\sigma J_\lambda^{\nu_1/n}). \qquad (3\cdot4)$$

Calculation of the posterior distributions and application of (3·2) results, except for a constant, in

$$K_A(\lambda \mid H, N) \propto \left\{1+\frac{\nu_2}{\nu_r}F(\lambda)\right\}^{-\frac{1}{2}(n-\nu_1)} (S_{\nu_r}^{(\lambda)}/J_\lambda^{2/n})^{-\frac{1}{2}\nu_2}. \qquad (3\cdot5)$$

The second term on the right-hand side causes the most important difference between (3·3) and (3·5). To appreciate the disadvantage of (3·5), recall that the posterior probability of λ under the general model with Box & Cox's prior is

$$p_2(\lambda \mid N, H, Y) \propto (S_{\nu_r}^{(\lambda)}/J_\lambda^{2/n})^{-\frac{1}{2}\nu_r}, \qquad (3\cdot6)$$

whence the second right-hand side term in (3·5) is $\{p_2(\lambda \mid N, H, Y)\}^{\nu_2/\nu_r}$, and is thus a monotonic function of $p_2(\lambda \mid N, H, Y)$. This term in (3·5) can be considered as a 'leakage' from the evidence of normality and homoscedasticity of a set of λ's, to the criterion for checking additivity.

In other words, if the model is nonadditive and the interactions are unremovable, the particular combination of $1/F$ and $p_2(\lambda \mid H, N, Y)$ in (3·5) will tend to make the region of coincidence of normality and homoscedasticity, and additivity overlap, leading to over-acceptance of the latter.

This 'leakage' phenomenon will be more important if ν_2/ν_r is far from zero. We ran the following simulation by analogy with the 'Textile example' analysed by Box & Cox, a single replicate of a 3^3 factorial experiment, where $\nu_2/\nu_r = 6/17$. We generated $\varepsilon_i \sim N(0,1)$ and constructed, for $i = 1, \ldots, 27$,

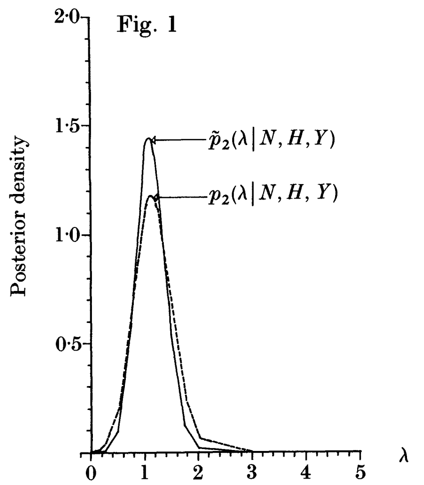
$$y_i = 1\cdot5 + 0\cdot16X_{1i} + 0\cdot08X_{2i} + 0\cdot1X_{3i} + 0\cdot115(X_{11i} + X_{12i} + X_{13i} + X_{22i} + X_{23i} + X_{33i}) + \varepsilon_i,$$

where X_{1i}, X_{2i} and X_{3i} are defined as in the 'Textile example', and $X_{11i} = X_{1i}X_{1i}$, $X_{12i} = X_{1i}X_{2i}$, etc.; see also Box & Tiao (1973, Chapter 10). Hence, the model is nonadditive and the true value of the transformation is $\lambda = 1$.

Figure 1 shows the posterior density of λ for the complex nonadditive model, for Box & Cox's prior and for the alternative. Both of them suggest the value $\lambda = 1$, the latter being more concentrated around its mode. The F ratios are shown in Fig. 3. The 5% significance point of the F statistic is 2·7, and the F value for $\lambda = 1$ is 2·77, lying outside the acceptance region, and for that value the additive model is rejected at 5% significance level from a classical point of view. A different situation occurs for Fig. 2, i.e. for the criteria for additivity when Box & Cox's priors are used. A maximum value is obtained at $\lambda = 0\cdot25$, and is bigger than one at $\lambda = 1$.

To overcome this difficulty and in the same spirit as in § 2, an alternative choice of non-outcome-dependent priors is proposed:

$$\tilde{p}_0(\theta, \sigma, \lambda \mid H, N) \propto 1/\sigma^{(k+1)}, \quad \tilde{p}_0(\theta_1, \sigma, \lambda \mid A, H, N) \propto 1/\sigma^{(\nu_1+1)}, \qquad (3\cdot7)$$



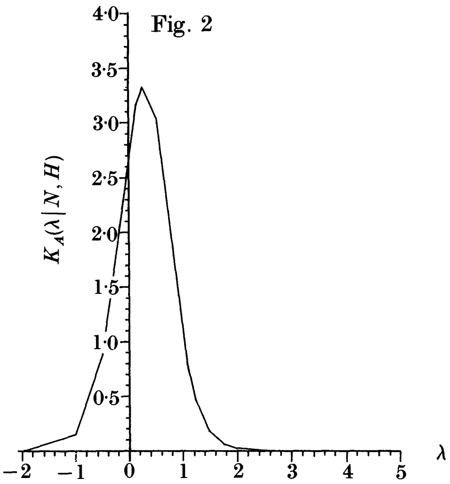
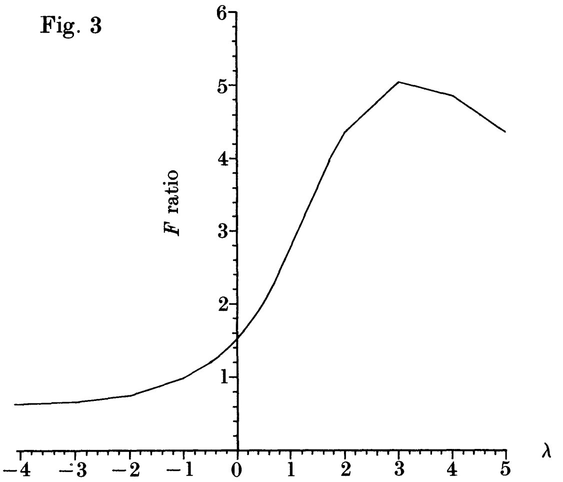


Fig. 1 (left). Posterior density for λ, assuming normality and homoscedasticity, broken line, when Box & Cox's prior (3·4) is used; $\hat{\lambda} = 1{\cdot}11$, $E(\lambda) = 1{\cdot}115$, $\sigma = 0{\cdot}337$; approximate 95% HPD interval $\{0{\cdot}45 \leqslant \lambda \leqslant 1{\cdot}77\}$. Solid line: when the alternative prior (3·7) is used; $\hat{\lambda} = 1{\cdot}11$, $E(\lambda) = 1{\cdot}11$, $\sigma = 0{\cdot}246$; approximate 95% HPD interval $\{0{\cdot}63 \leqslant \lambda \leqslant 1{\cdot}59\}$.

Fig. 2 (right). Ratio of posterior densities of the additive over the nonadditive model when Box & Cox's priors (3·4) are assumed.



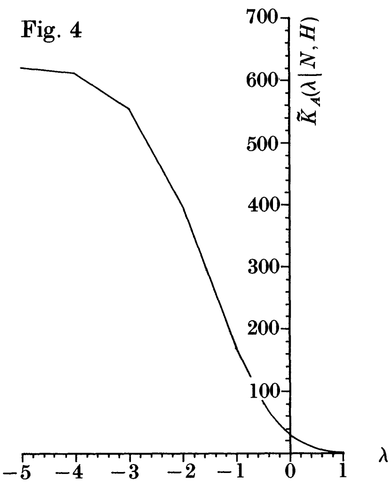


Fig. 3 (left). Values of F ratio. Fig. 4 (right). Ratio of posterior densities for λ of the additive model over the nonadditive model when the alternative priors (3·7) are assumed.

instead of (3·4). Integration out of θ and σ in the simpler and complex models and use of (3·2) lead to

$$\tilde{K}_A(\lambda\,|\,H, N) \propto \left\{1 + \frac{\nu_2}{\nu_r} F(\lambda)\right\}^{-\frac{1}{2}n}, \tag{3·8}$$

which is monotonic to the output of the non-Bayesian analysis for additivity, and does not show the leakage factor present in (3·5).

For the numerical example, $\tilde{K}_A(\lambda\,|\,N, H)$ is shown in Fig. 4. If the alternative priors (3·7) are employed, we get a monotonic function of the F statistic, decreasing with λ until $\lambda = 3$ and with a value less than one, 0·94, for $\lambda = 1$. Following the comments made at the end of

§ 3·1, note that in (3·4) and (3·7) neither of the right-hand sides are the priors in the left-hand sides conditioned on $\theta_2 = 0$. A similar observation will be valid for the next section.

3·3. *Criteria for homoscedasticity*

Suppose that there are K groups of data, expectation and variance being constant within each group. For the lth group, let σ_l^2 and $S^{(l)}$ be the variance and the residual sum of squares with $\nu_l = n_l - 1$ degrees of freedom and let $n = \Sigma n_l$, $S_\nu = \Sigma S^{(l)}$. The assumption at this point is just that there exists a transformation that induces normality simultaneously in all groups.

From the likelihood approach, and using again decomposition (3·1), the statistic for testing homoscedasticity in terms of the normalized variable $Z^{(\lambda)} = Y^{(\lambda)}/J_\lambda^{1/n}$ is

$$\{L_{\max}(\lambda \mid H, N) - L_{\max}(\lambda \mid N)\} = \log\left(\left[\prod_{l=1}^{k} \{S^{(l)}(\lambda, Z)/n\}^{\frac{1}{2}n_l}\right] \Big/ \{S_\nu(\lambda, Z)/n\}^{\frac{1}{2}n}\right), \qquad (3·9)$$

which is the log of the Neyman–Pearson L_1 criterion for testing H: $\sigma_1^2 = \ldots = \sigma_k^2$. At this point Box & Cox's choice of priors is

$$p_0(\theta, \sigma_1, \ldots, \sigma_k \mid N) \propto \prod_{l=1}^{k} \sigma_l^{-1}/J_\lambda^{k/n}, \quad p_0(\theta, \sigma, \lambda \mid N, H) \propto \sigma^{-1}/J_\lambda^{k/n}. \qquad (3·10)$$

The posterior distribution that results is in terms of the normalized Z's

$$K_H(\lambda \mid N) \propto \frac{\Pi\{S^{(l)}(\lambda, Z)\}^{\frac{1}{2}(n_l - 1)}}{\{S_\nu(\lambda, Z)\}^{\frac{1}{2}(n-k)}} \qquad (3·11)$$

and is equivalent to Bartlett's modification of L_1 statistics replacing sample sizes by degrees of freedom. On the other hand, instead of (3·10) we select the non-outcome-dependent priors, in the same spirit as the previous sections,

$$\tilde{p}_0(\theta, \sigma_1, \ldots, \sigma_k, \lambda \mid N) \propto \prod_{l=1}^{k} \sigma_l^{-2}, \quad \tilde{p}_0(\theta, \sigma, \lambda \mid N, H) \propto \sigma^{-(k+1)}.$$

Calculating the posteriors $\tilde{p}_2(\lambda \mid H, N, Y)$ and $\tilde{p}_2(\lambda \mid N, Y)$, and using the general formula (3·2), we get in terms of the Z's, except for a constant,

$$\tilde{K}_H(\lambda \mid N) \propto \left[\prod_{l=1}^{k} \{S^{(l)}(\lambda, Z)\}^{\frac{1}{2}n_l}\right] \Big/ \{S_\nu(\lambda, Z)\}^{\frac{1}{2}n},$$

which is the L_1 Neyman–Pearson criterion, and once again the advocated choice of prior (2·7) produces via Bayes's theorem a monotonic function of the statistic yielded by the likelihood approach to the problem.

4. A RELATED PROBLEM

Among other fields estimated transformations have had considerable impact in econometrics. Zellner (1971, Chapter 6) analysed the problem of generalized production functions,

$$\log V_i + \lambda V_i = \theta_1 + \theta_2 \log K_i + \theta_3 \log L_i + u_i, \qquad (4·1)$$

where the u_i's are independent normal errors with mean 0 and common variance σ^2, V is the economic output, K the capital, and L the labour. Zellner also carried out in parallel a likelihood and a Bayesian argument, arriving at different statistics. Following an argument similar to Box and Cox's, Zellner set up the 'outcome-dependent' prior

$$p_0(\theta_1, \theta_2, \theta_3, \sigma, \lambda) \propto 1/(\sigma J_\lambda^{3/n}). \tag{4·2}$$

If instead of (4·2), the non-outcome-dependent prior,

$$\tilde{p}_0(\theta_1, \theta_2, \theta_3, \sigma, \lambda) \propto 1/\sigma^4, \tag{4·3}$$

is chosen, then Zellner's likelihood and Bayesian solution for the posterior density of λ will coincide.

5. Conclusions

The alternative choice of priors advocated here appears to have advantages over the previous selection of priors. It removes the necessity of using an 'outcome-dependent' prior and makes the likelihood and Bayesian inference coincident, producing three appealing statistics

(a) the Neyman–Pearson likelihood ratio for the posterior odds about different λ's;
(b) the F ratio for testing additivity;
(c) the L_1 Neyman–Pearson statistic for testing homoscedasticity.

I am greatly indebted to Dr A. C. Atkinson, who guided this work, which was done at the Department of Mathematics, Imperial College, London. Also, I would like to thank Professor P. Bickel for a motivating conversation and Professor I. Rodriguez-Iturbe for his continuous encouragement. The author held a scholarship from CONICIT, Venezuela, during the research.

References

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations (with discussion). *J. R. Statist. Soc.* B **26**, 211–52.

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

Cox, D. R. (1977). The role of significance tests. *Scand. J. Statist.* **4**, 49–70.

Fraser, D. A. S. (1967). Data transformations and the linear model. *Ann. Math. Statist.* **38**, 1456–65.

Fraser, D. A. S. (1968). *The Structure of Inference.* New York: Wiley.

Lindley, D. V. (1961). The use of prior probability distributions in statistical inference. *Proc. 4th Berkeley Symp.* **1**, 453–68.

Lindley, D. V. (1972). *Bayesian Statistics, a Review.* SIAM, Philadelphia, Regional Conference Series in Applied Mathematics, 2. Philadelphia: Society for Industrial and Applied Mathematics.

Nelder, J. A. (1964). Contribution to discussion of paper by G. E. P. Box and D. R. Cox. *J. R. Statist. Soc.* B **26**, 244–52.

Raiffa, H. & Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Boston: Mass: Division of Research Graduate School of Business Administration, Harvard University.

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

[*Received December* 1979. *Revised June* 1980]