A Bayesian Approach to Statistical Inference about Climate Change

ANDREW R. SOLOW

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts*

(Manuscript received 9 July 1987, in final form 10 February 1988)

ABSTRACT

A Bayesian approach to statistical inference about climate change based on the two-phase regression model is presented. This approach is useful when nonobservational information is available about possible climate change. This information may refer to the timing or the nature of the possible change. The approach is applied to a historic temperature record.

1. Introduction

Solow 1987 used a classical analysis of the two-phase regression model to test for climate change in a historical record of hemispheric temperature. The purpose of this paper is to describe and apply a Bayesian analysis of the two-phase regression model. A Bayesian approach is useful when scientific information about climate change is available that is not in the form of measurements. This information may refer to both the timing and nature of the change.

There is substantial literature on problems of climate change. The comprehensive review article by Ellsaesser et al. (1986) contains many important references. Statistical work in this area has been carried out by Wigley and Jones (1981), Epstein (1982), Angell and Korshover (1983), and others. There is also substantial literature on Bayesian statistics. The monograph by Epstein (1985) focuses on climatological applications, and is clearly written and accessible. A more comprehensive theoretical account is given by Berger (1985). Leamer (1978) discusses Bayesian methods in linear regression models. Smith (1975) describes a general Bayesian approach to changepoint problems. Smith and Cook (1980) discuss a Bayesian approach to the two-phase regression model, although we do not use their approach here.

In section 2, we review the two-phase regression model and we present a general Bayesian analysis of the model. In section 3, we discuss our choice of prior distribution for some of the parameters in the model. The Bayesian approach is applied to a historic temperature record in section 4 and we make some concluding remarks in section 5.

* WHOI Contribution No. 6709.

Corresponding author address: Dr. Andrew Solow, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

2. A Bayesian analysis of the two-phase regression model

Suppose that we observe an annual temperature series, T_i, $i = 1, \cdots, n$. We will use a two-phase linear regression model to detect and characterize climate change. This model is written:

$$T_i = \begin{cases} a_0 + b_0 i + e_i, & i = 1, \cdots, r \\ a_1 + b_1 i + e_i, & i = r+1, \cdots, n \end{cases} \quad (1)$$

where the e_i form an independent sequence of normal noise with mean zero and unknown variance σ^2. The abscissa of the intersection of the two regression lines is

$$c = (a_0 - a_1)/(b_1 - b_0).$$

Although we will focus on the parameter c, which is called the changepoint, we note that r is also an unknown parameter. To ensure continuity of the underlying signal, we require that c lie in the interval $(r, r + 1)$. Without this constraint, the two-phase regression model may include a discontinuity in the signal at c.

Note that under (1), the prechange climate need not be stationary (i.e., b_0 need not be zero). It is possible to constrain this model to require prechange (or postchange) stationarity. However, in the absence of compelling external information, this is inadvisable, as it may give quite misleading results in the event of model misspecification.

Hinkley (1969, 1971) presents a classical (i.e., non-Bayesian) analysis of the two-phase regression model. The mathematical development is contained in the first of these papers, while a more general overview is contained in the second (which includes several typographical errors). Hinkley discusses maximum likelihood estimation of c and the two-sided likelihood ratio test of the null hypothesis $b_0 = b_1$. The distributional theory connected with the classical analysis of the two-phase regression model is quite difficult. For this reason, only asymptotic results are available, and even these are based on extensive simulation studies.

It is convenient to rewrite (1) as

$$T_i = a_0 + b_0 i + b(i - c)\mathrm{IND}_c(i) + e_i \qquad (2)$$

where

$$\mathrm{IND}_c(i) = \begin{cases} 0 & \text{if } i \leq c \\ 1 & \text{if } i > c \end{cases}$$

and where $b = b_1 - b_0$. For fixed c, (2) is a normal linear regression model with regressor variables i and $(i - c)\mathrm{IND}_c(i)$.

It is useful to introduce some matrix notation at this point. For fixed c, we will let

$$\mathbf{X}_c^t = \begin{cases} 1 & 1 \cdots 1 & 1 \cdots 1 \\ 1 & 2 \cdots r & r+1 \cdots n \\ 0 & 0 \cdots 0 & 1 \cdots n-r \end{cases}$$

$$\mathbf{T} = (T_1\, T_2 \cdots T_n)^t$$

$$\mathbf{a} = (a_0 b_0 b)^t$$

$$\mathbf{e} = (e_1\, e_2 \cdots e_n)^t$$

$$\mathbf{a}_c^* = (\mathbf{X}_c^t\mathbf{X}_c)^{-1}\mathbf{X}_c^t\mathbf{T}$$

$$\mathrm{RSS}_c = (\mathbf{T} - \mathbf{X}_c\mathbf{a}_c^*)^t(\mathbf{T} - \mathbf{X}_c\mathbf{a}_c^*)$$

where the superscript t denotes transpose. In terms of this notation, (2) can be written

$$\mathbf{T} = \mathbf{X}_c\mathbf{a} + \mathbf{e}$$

while a_c^* is the ordinary least-squares estimate of a, and RSS_c is the corresponding residual sum of squares.

Adopting the Bayesian viewpoint, we treat the changepoint as a random variable C and we assume that a prior distribution, $\pi(c)$, is specified. The posterior distribution of C, given T, is given by Bayes' theorem:

$$p(c/\mathbf{T}) = p(\mathbf{T}/c)\pi(c) \Big/ \int p(\mathbf{T}/c)\pi(c)dc \qquad (3)$$

where $p(c/\mathbf{T})$ is the posterior distribution of C, given $\mathbf{T}$, and $p(\mathbf{T}/c)$ is the likelihood of $\mathbf{T}$, given $C = c$, and where the integral in the denominator of (3) is taken over the parameter space of C.

We also treat the other model parameters (a_0, b_0, b, and σ^2) as random variables. Let Θ denote the vector random variable of these parameters, and let $A = (A_0 B_0 B)^t$ denote the first three elements of Θ. We will refer to $\mathbf{A}$ as the regression parameters, and we will use $\mathbf{a}$ to denote a realization of $\mathbf{A}$. The likelihood in (3) can be expressed as

$$p(\mathbf{T}/c) = \int p(\mathbf{T}/c, \Theta)\pi(\Theta/c)d\Theta \qquad (4)$$

where $p(\mathbf{T}/c, \Theta)$ is the likelihood of $\mathbf{T}$ given $C = c$ and $\Theta = \Theta$, $\pi(\Theta/c)$ is the conditional prior distribution of Θ given $C = c$, and the integral in (4) is taken over the parameter space of Θ.

Once the joint prior distribution of the parameters C and Θ,

$$\pi(c, \Theta) = \pi(c)\pi(\Theta/c),$$

is specified and the observations, $\mathbf{T}$, are made, the posterior distribution of C can be found from (3) and (4). This posterior distribution provides the basis for Bayesian inference about C. An estimate of C can be obtained, for example, by taking the posterior mean (or, more generally, by minimizing the expected loss for some appropriate choice of loss function). The Bayesian analogue of a confidence interval, which is called a credible set, can also be found. From the Bayesian viewpoint, since $p(c/\mathbf{T})$ is an actual probability distribution for C, we can speak meaningfully of the probability that C lies within a certain set. This is in contrast to classical confidence procedures, which can only be interpreted in terms of coverage probability (Berger 1985). A credible set for C of size $1 - \alpha$ is a subset, S_α, of the parameter space of C such that

$$\int_{S_\alpha} p(c/\mathbf{T})dc = 1 - \alpha.$$

In choosing a credible set of fixed size, it is sensible to include only those points with highest posterior density. A credible set for C of size $1 - \alpha$ with highest posterior density, H_α, is the $1 - \alpha$ credible set such that the posterior density of each point in H_α is at least as high as the posterior density of each point outside H_α. We will consider specific examples of these procedures in section 4.

The posterior distribution of Θ can also be found from

$$p(\Theta/\mathbf{T}) = \int p(\Theta/\mathbf{T}, c)p(c/\mathbf{T})dc \qquad (5)$$

where the integral is taken over the interval $(1, n)$, and where $p(\Theta/\mathbf{T}, c)$ is the conditional posterior distribution of Θ given $C = c$. By Bayes' theorem:

$$p(\Theta/\mathbf{T}, c) = p(\mathbf{T}/c, \Theta)\pi(\Theta/c)/p(\mathbf{T}/c).$$

We will be particularly interested in the marginal posterior distribution of B, which can be found by integrating $p(\Theta/\mathbf{T})$ over a_0, b_0, and σ^2. This marginal posterior distribution, $p(b/\mathbf{T})$, provides the basis for Bayesian inference about B. A test for climate change can be carried out by considering the null hypothesis H_0: $B = 0$. A Bayesian test for this hypothesis can be based on the posterior odds ratio:

$$\mathrm{prob}(H_0/\mathbf{T})/[1 - \mathrm{prob}(H_0/\mathbf{T})].$$

This ratio can be expressed as

$$[p(\mathbf{T}/H_0)/p(\mathbf{T}/H_1)][\pi(H_0)/(1 - \pi(H_0))] \qquad (6)$$

where H_1 is the complement of H_0 (i.e., $B > 0$), $p(\mathbf{T}/H_i)$ is the likelihood of $\mathbf{T}$ under H_i, and $\pi(H_0)$ is the prior probability attached to H_0. In the context of testing a point null hypothesis, the prior distribution must attach a mass of probability to H_0; otherwise, $p(H_0/\mathbf{T})$ will be identically zero. The likelihood $p(\mathbf{T}/H_0)$—which is independent of C—is given by

$$p(\mathbf{T}/H_0) = \int p(\mathbf{T}/\Theta)\pi(\Theta/H_0)d\Theta \qquad (7)$$

where the integral is taken over the subspace of the parameter space of Θ satisfying $b = 0$, and $\pi(\Theta/H_0)$ is the prior distribution of Θ under H_0. The likelihood $p(T/H_1)$ does depend on C and is given by

$$p(\mathbf{T}/H_1) = \int p(\mathbf{T}/H_1, c)\pi(c/H_1)dc \qquad (8)$$

where $\pi(c/H_1)$ is the prior distribution of C under H_1. The likelihood $p(T/H_1, c)$ is given by

$$p(\mathbf{T}/H_1, c) = \int p(\mathbf{T}/\Theta, c)\pi(\Theta/H_1, c)d\Theta \qquad (9)$$

where $\pi(\Theta/H_1, c)$ is the conditional prior distribution of Θ under H_1 given $C = c$ and where the integral is taken over the subspace of the parameter space of Θ with $b > 0$. We note in passing that this test for climate change is equivalent to testing the null hypothesis H_0: $C = n$. Again, we will consider specific examples in section 4.

In section 4, we will apply the Bayesian methods outlined in this section to an historic temperature record. To keep the discussion focused, we will choose a prior distribution for Θ that is analytically tractable, and that is independent of C. This prior distribution is discussed in section 3. Analytical tractability is extremely convenient, particularly in an expository context. In practice, however, the choice of prior distribution should reflect actual prior information, and numerical methods should be used if analytical solutions are difficult. The assumption of prior independence of Θ and C may also be unreasonable in practice. For example, we may believe that a change occurring late in the record will be more rapid than a change occurring early in the record. Again, while the assumption of prior independence is convenient, prior dependence can be readily handled in practice.

3. Specification of $\pi(\Theta)$

A convenient choice for $\pi(\Theta)$ is the normal-gamma distribution:

$$\pi(\Theta) = \pi(a/\sigma^{-2})\pi(\sigma^{-2})$$

where we have parameterized in terms of σ^{-2}, which is called the precision. Under the normal-gamma prior, the conditional distribution of the regression parameters given the precision is trivariate normal with mean vector $\mathbf{a}$ and variance matrix $\sigma^2\mathbf{N}$. Also, the unconditional prior distribution of the precision is gamma with location parameter s^2 and scale parameter v (i.e., with mean $1/s^2$ and variance $2/vs^4$).

For our purposes, the usefulness of this prior distribution stems from the following three results (Leamer 1978).

First, if Θ has a normal-gamma distribution with parameters $\mathbf{a}$, $\mathbf{N}$, s^2, and v, then the marginal (i.e., unconditional) distribution of the regression parameters is trivariate Student with parameters a, $s^2\mathbf{N}$, and v (i.e., with mean vector a and variance matrix $s^2\mathbf{N}(v/(v-2))$.

Second, if Θ has the normal-gamma prior distribution with parameters $\mathbf{a}$, $\mathbf{N}$, s^2, and v, then the conditional posterior distribution of Θ given $C = c$ is also normal-gamma with parameters:

$$\mathbf{a}_c' = (\mathbf{N}^{-1} + \mathbf{X}_c'\mathbf{X}_c)^{-1}(\mathbf{N}^{-1}\mathbf{a} + \mathbf{X}_c'\mathbf{X}_c\mathbf{a}_c^*) \qquad (10a)$$

$$\mathbf{N}_c' = (\mathbf{N}^{-1} + \mathbf{X}_c'\mathbf{X}_c)^{-1} \qquad (10b)$$

$$s_c'^2 = (v + n)^{-1}(v\mathbf{s}^2 + \mathrm{RSS}_c + (\mathbf{a}_c^* - \mathbf{a})'\mathbf{N}^{-1} \times \mathbf{N}_c'(\mathbf{X}_c'\mathbf{X}_c)(a_c^* - \mathbf{a}) \qquad (10c)$$

$$v_c' = v + n. \qquad (10d)$$

Third, if we let

$$k = v^{v/2}\Gamma((v + n)/2)/\pi^{n/2}\Gamma(v/2) \qquad (11a)$$

$$\mathbf{M}_c = \mathbf{I}_n - \mathbf{X}_c(\mathbf{N}^{-1} + \mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c' \qquad (11b)$$

$$Q_c = (\mathbf{T} - \mathbf{X}_c\mathbf{a})'(\mathbf{T} - \mathbf{X}_c\mathbf{a}) - (\mathbf{a}_c^* - \mathbf{a})' \times (\mathbf{X}_c'\mathbf{X}_c)\mathbf{N}'(\mathbf{X}_c'\mathbf{X}_c)(\mathbf{a}_c^* - \mathbf{a}) \qquad (11c)$$

where Γ is the gamma function and $\mathbf{I}_n$ is the n-by-n identity matrix, then the likelihood in (4) is given by the n-variate Student function:

$$p(\mathbf{T}/c) = k|\mathbf{M}_c/s^2|^{1/2}[v + (Q_c/s^2)]^{-(v+n)/2}. \qquad (12)$$

We are now in a position to perform the Bayesian analysis outlined in section 2. The posterior distribution of C can be found from (3) using the likelihood given in (12). In practice, we would discretize the interval $(1, n)$, replace the continuous probability density functions $\pi(c)$ and $p(c/\mathbf{T})$ by probability mass functions defined at the discretization points, and replace the integral in (3) by a discrete sum. The posterior distribution of Θ can be found from (5), using the fact that the conditional posterior distribution of Θ, given $C = c$, is normal-gamma with parameters given by (10). Again, in practice, we would discretize the interval $(1, n)$ and replace the integral in (5) by a discrete sum. To find the posterior distribution of B, we replace $p(\Theta/\mathbf{T}, c)$ by $p(b/\mathbf{T}, c)$—which is Student with parameters given by the appropriate elements of $\mathbf{a}_c'$, $s_c'^2\mathbf{N}_c'$, and v_c'—in (5).

Although the main advantage of the Bayesian approach arises when specific prior information is available, it is sometimes the case that no such information is available about some of the parameters. In that case, one approach is to specify a so-called vague prior distribution for those parameters. The member of the normal-gamma family that is used to reflect prior ignorance about Θ has parameters $\mathbf{N}^{-1} = 0$ and $v = 0$. A difficulty arises in finding the likelihood $p(\mathbf{T}/c)$ for this choice of prior parameters. The problem which is discussed by Leamer (1978), is that the expression for this likelihood depends on the way in which $\mathbf{N}^{-1}$ becomes small. The option that we will choose leads to

$$p(\mathbf{T}/c) \propto (\mathrm{RSS}_c)^{-n/2} \tag{13}$$

with the constant of proportionality independent of c. We prefer this choice for the following reason. If $\pi(c)$ is uniform over the interval $(1, n)$—which is a vague prior for C—then the value of C that maximizes posterior likelihood is identical to the classical maximum likelihood estimate. This follows from the fact that the classical maximum likelihood estimate is found by minimizing RSS_c. This agreement of Bayesian and classical results in the case of prior ignorance is appealing.

The parameters of the conditional posterior distribution for Θ given $C = c$ corresponding to the vague prior do not depend on the way in which $\mathbf{N}^{-1}$ becomes small. From (7), these are given by

$$\mathbf{a}'_c = a^*_c \tag{14a}$$

$$\mathbf{N}'_c = (\mathbf{X}_c{}'\mathbf{X}_c)^{-1} \tag{14b}$$

$$v'_c = n \tag{14c}$$

$$s'^2_c = \mathrm{RSS}_c/n. \tag{14d}$$

The marginal conditional posterior distribution of the regression parameters is trivariate Student with pa rameters $\mathbf{a}^*_c$, $(\mathrm{RSS}_c/n)(\mathbf{X}_c{}'\mathbf{X}_c)^{-1}$, and n. This reproduces the classical result except that the degrees-of-freedom parameter is n instead of $n - 3$.

We note in passing that the normal-gamma prior distribution assumes prior dependence between the regression parameters and the precision. This may or may not be reasonable in practice. One argument in favor of prior dependence is that if we discover that the precision is low (i.e., the process is noisy), we may wish to increase our prior variance for the regression parameters. An alternative prior distribution that assumes prior independence is discussed by Dickey (1975). This alternative prior distribution is more difficult to handle analytically.

4. Application

In this section, we apply the Bayesian procedures described in section 2 to an annual series of Southern Hemisphere temperature deviations compiled by Jones et al. (1986) and updated by Jones (1985). This is the same dataset analyzed by Solow (1987). Because a constant has been subtracted from each observation, inference about a_0 is meaningless. The observations, which run from 1858 to 1985, are plotted in Fig. 1. The way in which this dataset—which is referred to as SH60—was constructed, and problems relating to its quality, are discussed by Jones et al. (1986).

Throughout this analysis, we will use a vague prior for Θ. In a more practical context, involving other climatological information, a more informative prior distribution would probably be appropriate. However, preliminary results suggest that the results of the analysis described below do not depend strongly on the specification of $\pi(\Theta)$, unless prior variances are quite small. The explanation for this is that n is large enough so that the influence of the data tends to dominate the influence of the prior in the posterior.

We will consider three prior distributions for the changepoint C:

$$\pi_1(c) = 1/119, \qquad 1862 \leqslant c \leqslant 1980$$

$$\pi_2(c) = \begin{cases} (0.25)1/60, & 1862 \leqslant c \leqslant 1921 \\ (0.75)1/59, & 1921 < c \leqslant 1980 \end{cases}$$

$$\pi_3(c) = \begin{cases} (0.25)1/89, & 1862 \leqslant c \leqslant 1950 \\ (0.75)1/30, & 1950 < c \leqslant 1980. \end{cases}$$

The first distribution reflects prior ignorance about C. The second distribution attaches prior probability of 0.25 to the event that C lies in the first half of the record and 0.75 to the event that C lies in the second half of the record, although within each half no further prior information is available. The third distribution further concentrates prior probability in the last 30 yr of the record. In all cases, we exclude the possibility that C lies in the first 4 or last 5 yr of the record.

The posterior distributions of C corresponding to these prior distributions are shown in Fig. 2. The prior distributions are also shown for reference. The general shapes of the posterior distributions are similar, with major peaks at 1887 and 1976. For $\pi_1(c)$ and $\pi_2(c)$, the posterior modes occur at 1887, while for $\pi_3(c)$ the posterior mode occurs at 1976. The highest posterior density credible set for C with size 0.50 for $\pi_1(c)$ is given by the interval (1879, 1899). For $\pi_2(c)$, $H_{0.5}$ consists of the disjoint intervals (1879, 1904) and (1973, 1978), containing posterior probabilities of 0.41 and 0.09, respectively. For $\pi_3(c)$, $H_{0.5}$ consists of the disjoint intervals (1884, 1893) and (1967, 1980), containing posterior probabilities of 0.15 and 0.35, respectively. A different way to characterize the posterior distribution of C is by the median. For $\pi_1(c)$, the median occurs around 1892. That is, there is an equal posterior probability that C occurs in the interval (1862, 1892) as in the interval (1893, 1980). For $\pi_2(c)$, the median occurs around 1902. For $\pi_3(c)$, the median occurs around 1951.

The posterior distributions of B_0 and B for the three prior distributions of C are shown in Figs. 3 and 4, with some summary statistics given in Tables 1 and 2. For reference, the ordinary least squares regression es-

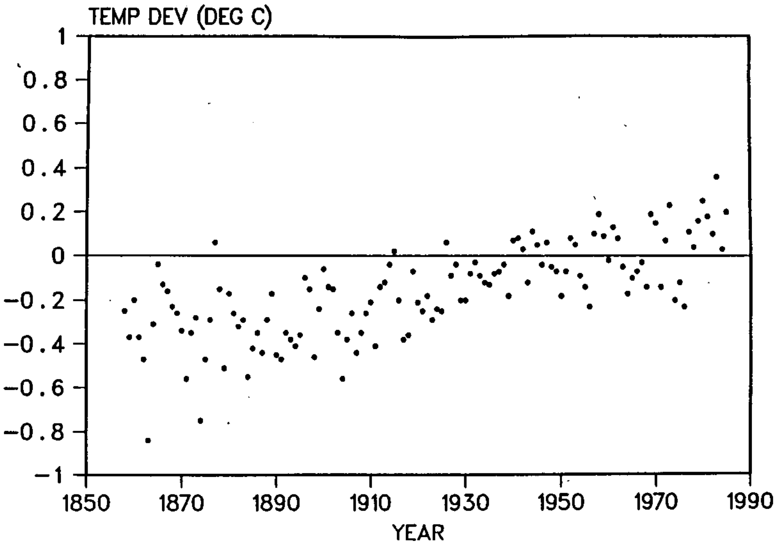

FIG. 1. Southern Hemisphere surface air temperature deviations, 1858–1985 [from Jones (1985), Jones et al. (1986)].

timates of B_0 and B for $C = 1887$ are $-0.00076°\text{C y}^{-1}$ and $0.0055°\text{C y}^{-1}$, respectively, while for $C = 1976$, these estimates are $0.0038°\text{C y}^{-1}$ and $0.0173°\text{C y}^{-1}$, respectively. Also, the ordinary least squares estimate of B_0 in the null (i.e., no change) model,

$$T_i = A_0 + B_0 i + e_i, \quad i = 1858, \cdots, 1985, \quad (15)$$

is $0.00403°\text{C y}^{-1}$.

The posterior distributions of B_0 are skewed to the left and each has a mode at around $0.004°\text{C y}^{-1}$. This mode becomes sharper as $\pi(c)$ concentrates more probability late in the record. This is explained in the following way. From (5), the posterior distribution of B_0 is a linear combination of conditional posterior distributions of B_0 given C. The weight function in this linear combination is the posterior density for C. The means and standard deviations of these conditional posterior distributions of B_0 are shown in Fig. 5. The conditional posterior distributions of B_0 for C occurring before circa 1880 have relatively low means and high standard deviations. It is these distributions that cause the skewness in the posterior distribution of B_0. On the other hand, the conditional posterior distributions of B_0 for C occurring after circa 1880 all have means around $0.004°\text{C y}^{-1}$ with relatively low standard deviations. It is these distributions that cause the peak in the posterior distribution of B_0. As the prior distribution for C concentrates probability late in the record, the weight on the first set of conditional posterior distributions decreases and the weight on the second set increases. This causes the skewness to become weaker and the peakedness to become stronger.

The posterior distributions for B are skewed to the right, with the skewness increasing somewhat as $\pi(c)$ concentrates more probability late in the record. The explanation for this is that the conditional posterior distributions of B for C occurring both early and late in the record have relatively high means and relatively high standard deviations. These are shown in Fig. 6. This causes the posterior distribution of B to be skewed to the right. As the prior distribution for C concentrates probability late in the record, this skewness becomes stronger. This effect would also occur if the prior distribution for C concentrated probability early in the record.

We can also find a convenient approximation to the posterior distribution of the post-change slope parameter, $B_1 = B_0 + B$, in the following way. Because the degrees-of-freedom parameters of the conditional posterior distributions of $\mathbf{A}$, given $C = c$, are all large, we can approximate these trivariate Student distributions by trivariate normal distributions with mean vectors a_c^* and variance matrices $(\text{RSS}_c/n)(\mathbf{X}_c'\mathbf{X}_c)^{-1}$. The conditional posterior distribution of B_1, given $C = c$—$p(b_1/\mathbf{T}, c)$—is therefore approximately normal with mean

$$b'_{1c} = b'_{0c} + b'_c$$

and variance

$$v'_{1c} = v'_{0c} + v'_c + 2\gamma'_c$$

where b'_{0c} and b'_c are the conditional posterior means of B_0 and B, v'_{0c} and v'_c are the conditional posterior variances of B_0 and B and γ'_c is the conditional posterior covariance between B_0 and B. Finally, the unconditional posterior distribution of B_1 is given by

$$p(b_1/\mathbf{T}) = \int p(b_1/\mathbf{T}, c)p(c/\mathbf{T})dc.$$

The posterior distributions of B_1 found in this way

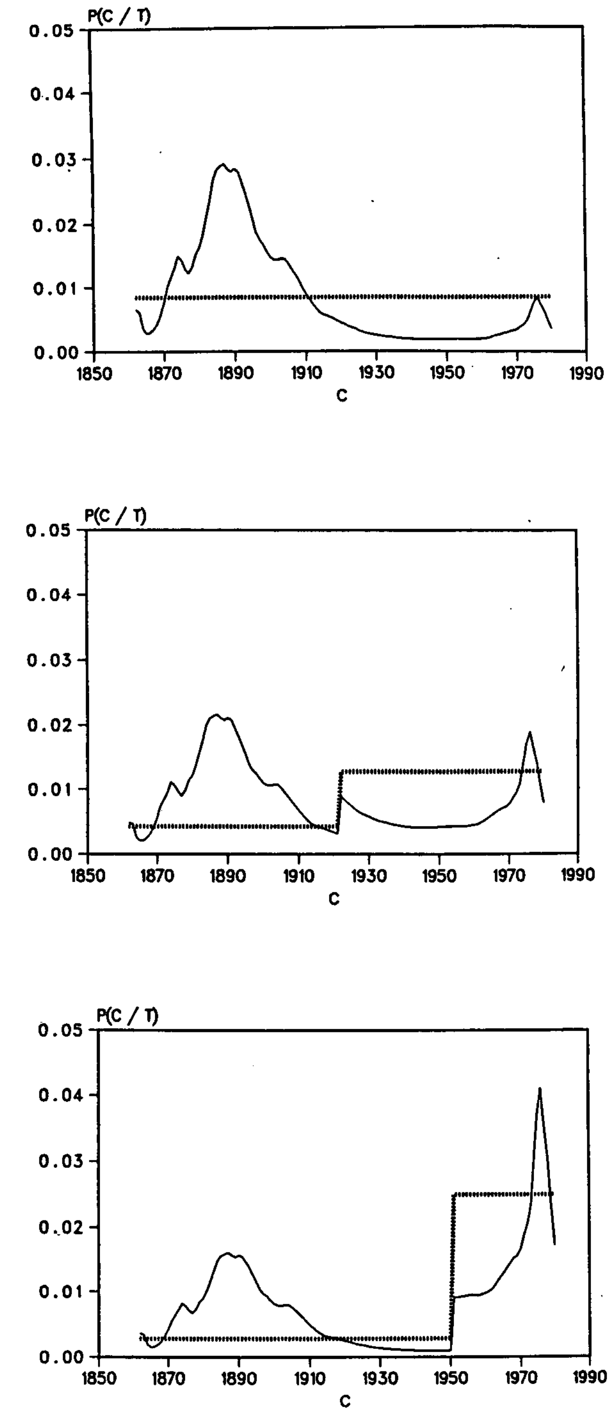

FIG. 2. Posterior distributions of C for three choices of prior distribution of C (shown by dashed line).

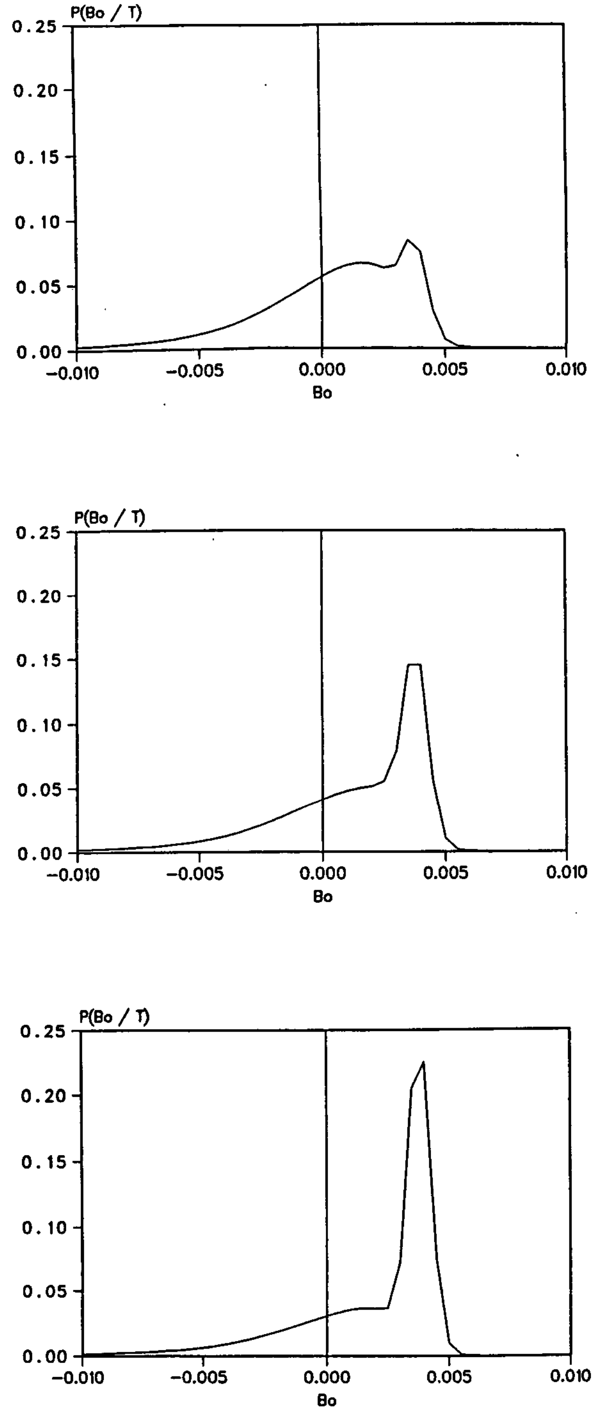

FIG. 3. Posterior distributions of B_0 for three choices of prior distribution of C.

for the three prior distributions of C are shown in Fig. 7, with some summary statistics given in Table 3. These posterior distributions are skewed to the right, with the skewness increasing as $\pi(c)$ concentrates more probability late in the record. They each show a mode at around 0.0045°C y^{-1}, with peakedness decreasing as skewness increases. The conditional posterior means and standard deviations of B_1 are shown in Fig. 8. The conditional posterior mean and standard deviation are relatively high for C occurring late in the record. This

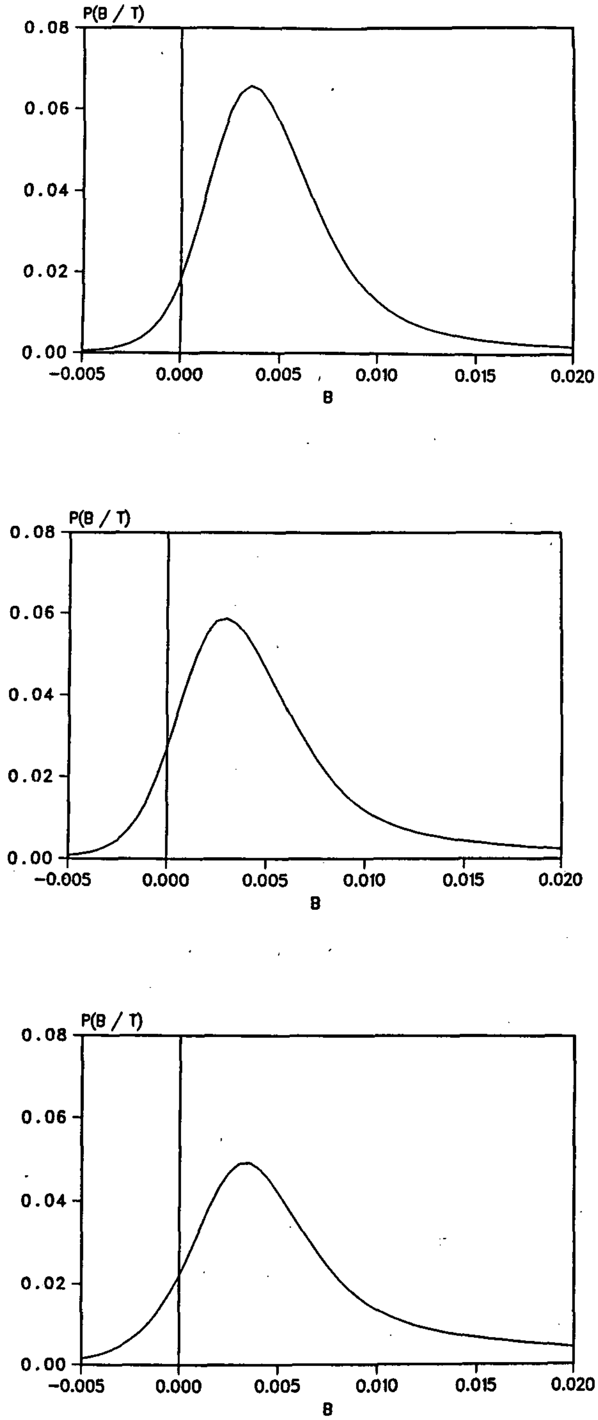

FIG. 4. As in Fig. 3, except for posterior distribution of B.

explains the skewness in the unconditional posterior distribution of B_1. For C occurring before circa 1950, the conditional posterior mean is around 0.0045°C y^{-1}, with relatively low conditional posterior standard deviation. This explains the peakedness in the unconditional posterior distribution of B_1. As the prior distribution of C concentrates more probability late in the record, the skewness increases and the peakedness decreases. The posterior correlation between B_0 and B is −0.78 for $\pi_1(c)$, −0.57 for $\pi_2(c)$, and −0.33 for $\pi_3(c)$.

TABLE 1. Summary statistics from the posterior distributions of B_0 for three choices of prior distribution of C.

Statistic	π_1	π_2	π_3
Mean	−0.00033	0.00072	0.00153
Standard deviation	0.00640	0.00571	0.00520
$H_{0.5}$	(0.00095, 0.0045)	(0.002, 0.005)	(0.003, 0.005)

We turn next to the problem of testing the null hypothesis of no change, H_0: $B = 0$. From (6), the posterior odds ratio for H_0 is given by the product of the likelihood ratio and the prior odds ratio. For our choice of vague prior distribution of Θ, the likelihood under H_0, $p(\mathbf{T}/H_0)$, is proportional to $(\mathrm{RSS})^{-64}$, where RSS is the residual sum of squares from fitting the null model (15). The likelihood under the alternative hypothesis H_1: $B \neq 0$ is given by (8) and (9). If, under H_1, we specify a vague normal-gamma prior distribution for Θ excluding the subspace with $b = 0$ (which has zero measure), then $p(\mathbf{T}/H_{1,c})$ is proportional to $(\mathrm{RSS}_c)^{-64}$. Finally, the likelihood ratios for $\pi_1(c)$, $\pi_2(c)$, and $\pi_3(c)$ are 0.202, 0.302, and 0.334, respectively. For any choice of $\pi(H_0)$, we can now find the corresponding posterior odds ratio. Alternatively, by simple algebra, we can find prob($H_0/\mathbf{T}$), which is the Bayesian analogue to the classical type I error probability. For example, if we attach equal prior probabilities to the null and alternative hypotheses (i.e., $\pi(H_0) = \pi(H_1) = 0.5$), the posterior probabilities of H_0 for the three choices of $\pi(c)$ are 0.168, 0.232, and 0.250, respectively. On the other hand, if we attach prior probabilities of 0.25 to the null hypothesis, and 0.75 to the alternative hypothesis, the posterior probabilities of H_0 for the three choices of $\pi(c)$ are 0.063, 0.091, and 0.100, respectively.

5. Discussion

From a practical point of view, the Bayesian approach to statistical inference provides a flexible, rigorous way to combine observational data with other kinds of information. In situations where there is no nonobservational information about some or all of the

TABLE 2. Summary statistics from the posterior distributions of B for three choices of prior distribution of C.

Statistic	π_1	π_2	π_3
Mean	0.00602	0.00615	0.00783
Standard deviation	0.00735	0.00818	0.00975
$H_{0.5}$	(0.0015, 0.0065)	(0.00095, 0.006)	(0.00095, 0.007)

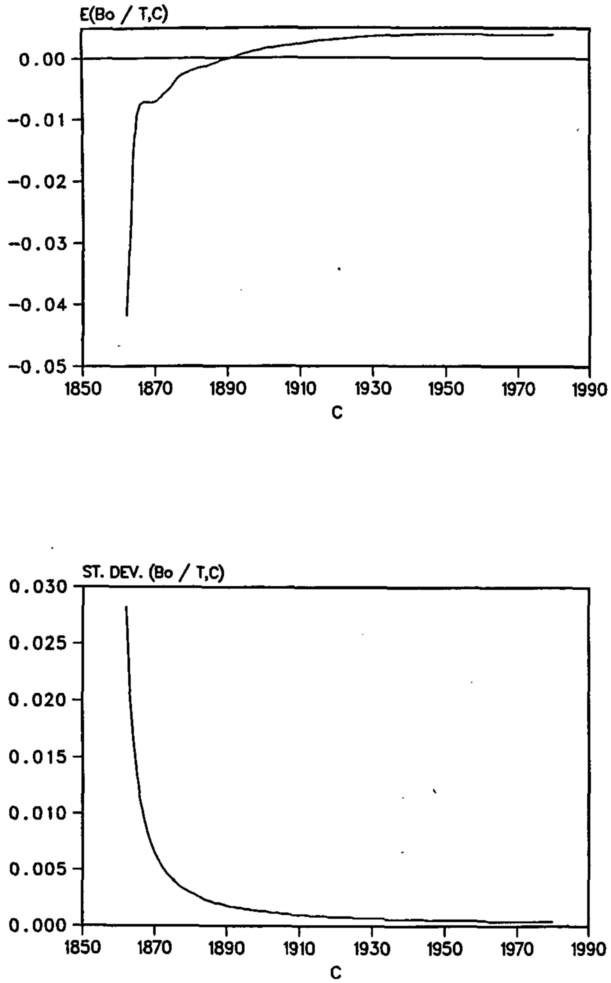

FIG. 5. Conditional posterior means and standard deviations of B_0 given C.

model parameters, the use of a vague prior distribution for those parameters allows the Bayesian approach to proceed, using a more informative prior distribution for the remaining parameters, if any. Another advantage of the Bayesian approach is that it provides results with clear (and, to many, intuitively appealing) interpretations. A good example of this arises in the context of hypothesis-testing. Under the Bayesian approach, we may speak meaningfully of the probability that a null hypothesis is true. If we reject the null hypothesis, we know that prob(H_0/**T**) is the probability that we have made an error. This is quite different from the classical type I error probability, which refers to an imaginary sequence of realizations of the model. In addition, because prob(H_0/**T**) is a true probability, the Bayesian approach allows us to choose a rejection rule in a systematic way. If we let $L(i/j)$ be the loss associated with accepting H_i when H_j is true, and if we let $L(i/i) = 0$, then the Bayes rule for choosing between H_0 and H_1 is to accept H_0 if and only if

$$\text{prob}(H_0/\mathbf{T})/\text{prob}(H_1/\mathbf{T}) > L(0/1)/L(1/0),$$

and otherwise to accept H_1. This rule minimizes posterior expected loss. Again, this is in contrast to classical hypothesis-testing, where there is little systematic guidance in choosing a critical level. In terms of testing for climate change, if we are concerned with early detection, we would tend to make $L(0/1)$ relatively small. On the other hand, if we are concerned with raising fears, we would tend to make $L(0/1)$ relatively large.

Before turning to the results of section 4, we make three remarks about the Bayesian approach. First, like the classical approach, inferences based on the Bayesian approach can be quite sensitive to model misspecification. For example, if the noise terms are not independently normally distributed with constant variance, the conditional likelihood $p(\mathbf{T}/c)$ will not be given by the multivariate Student function (12). The use of a Bayesian approach does not absolve us of carefully modeling the data. Second, in applying the Bayesian approach, it is important that the specification of prior

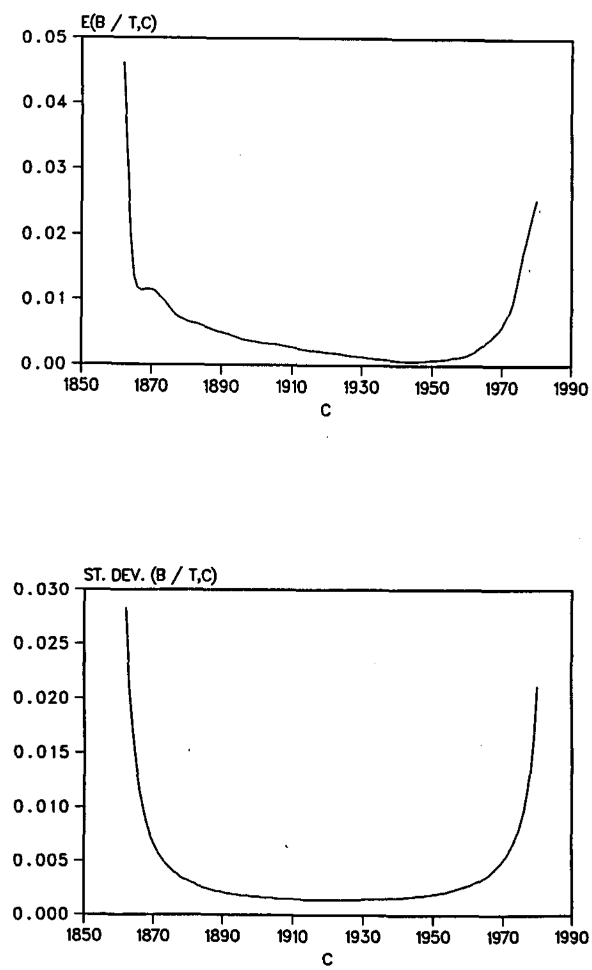

FIG. 6. As in Fig. 5, except for B given C.

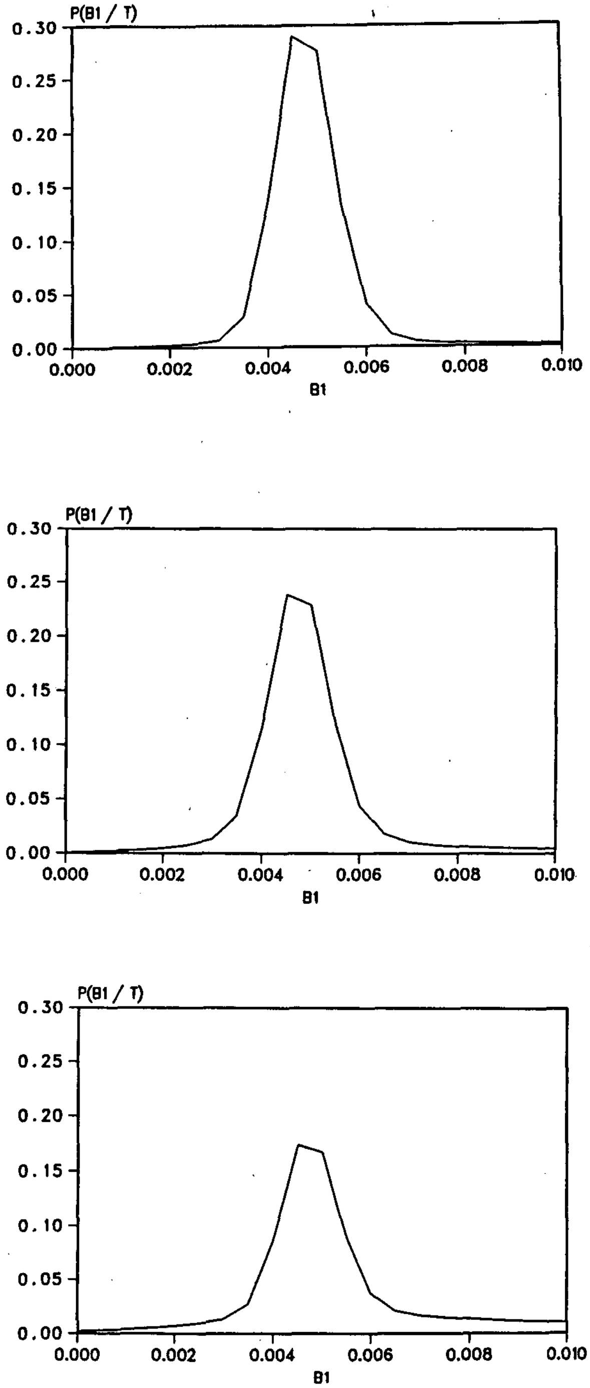

FIG. 7. As in Fig. 3, except for posterior distribution of B_1.

distributions be made without reference to the data with which they will be updated. If the distinction between prior information and observational information is not maintained, the prior distribution will already be partially posterior, and inferences will be spuriously sharp. This is the Bayesian counterpart to data-dredging. Third, in the interests of clarity, we have made

TABLE 3. Summary statistics from the posterior distributions of B_1 for three choices of prior distribution of C.

Statistic	π_1	π_2	π_3
Mean	0.00570	0.00687	0.00936
Standard deviation	0.00469	0.00678	0.00943
$H_{0.5}$	(0.004, 0.0055)	(0.004, 0.0055)	(0.0035, 0.006)

certain prior assumptions (which were noted at the time) that may be unrealistic in practice. While a more realistic Bayesian analysis of climate data is likely to be analytically messier, numerical methods allow us considerable latitude in model choice and specification of prior distributions.

We turn briefly to the results of section 4. Perhaps the most striking result is the skewness of the posterior distributions of B_0, B, and B_1. As we have seen, this skewness is caused by the combination of anomalous behavior of the conditional posterior means and high

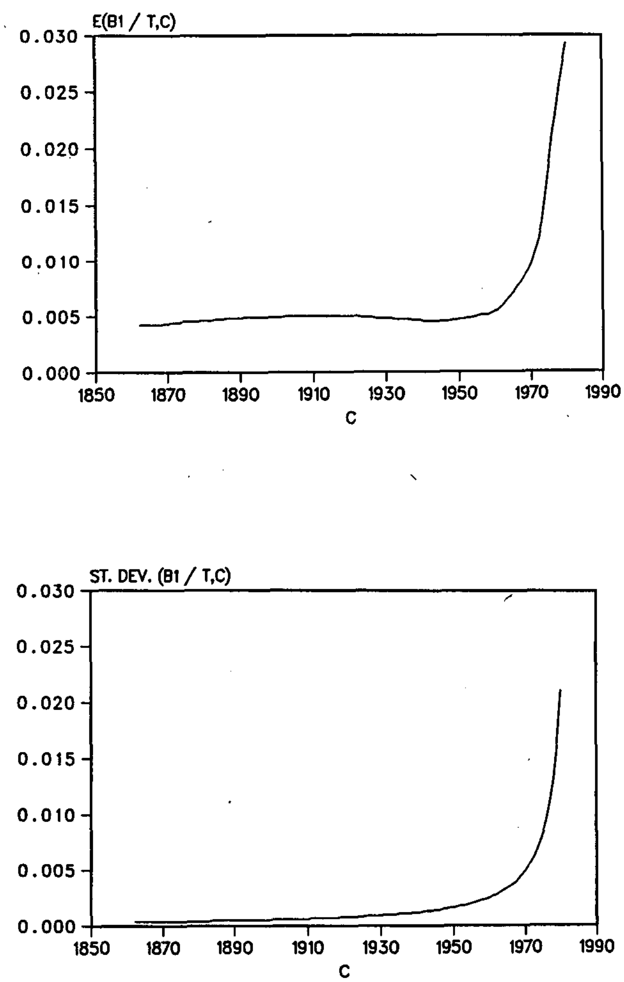

FIG. 8. As in Fig. 5, except for B_1 given C.

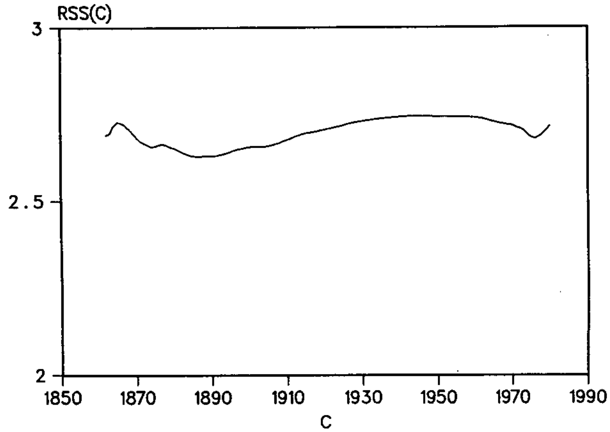

FIG. 9. Residual sum of squares from fitting model (2) for fixed C.

posterior standard deviations of B_0, B, and B_1 at the beginning and end of the record. The high standard deviations are primarily due to the design effect (i.e., the fact that there are only a few observations available to estimate the prechange (postchange) parameters near the beginning (end) of the record), and not to any large fluctuations in RSS_c, which is shown in Fig. 9. This causes the diagonal elements of $(\mathbf{X}_c{}^t\mathbf{X}_c)^{-1}$ to be high at the ends of the record. One consequence of this is that the posterior odds in favor of climate change are actually lower as we concentrate more prior probability for the changepoint late in the record, despite the fact that the conditional posterior mean of B is higher late in the record. The behavior of the means at the ends of the record is not an end-effect, although it is difficult to judge their significance because of the high standard deviations. For these reasons, inference about climate change becomes more difficult as we focus more on the later part of the record. To some extent, the use of a more informative prior distribution for the model parameters would help sharpen this inference. However, since much of what is known about the historic climate is based on datasets such as SH60, truly external information may not be available to refine prior distributions.

In addition to using more informative prior distributions, it is possible to extend the Bayesian analysis in a number of ways. One interesting extension would be to consider a three-phase regression model. As mentioned by Solow (1987), if we attach zero prior probability to the event that the two changepoints in such a model lie within 10 years of each other, and if we use vague prior distributions for both the changepoints and the model parameters, the pair of changepoints with highest posterior density is (1888, 1978). Under the Bayesian approach, it would be straightforward to provide a credible set for the two changepoints, or to test the hypothesis that there is only one changepoint against the alternative that there are two. This would not be the case under the classical approach, where the distributional results are not available. Finally, it is possible in principle to further extend the Bayesian approach to include the number of changepoints as an unknown parameter. However, the specification of the joint prior distribution of the model parameters would be complex in this case.

Acknowledgments. This work was supported by the J. N. Pew Charitable Trusts and the Marine Policy Center of the Woods Hole Oceanographic Institution. The author gratefully acknowledges (yet again) the comments of Dr. E. S. Epstein.

REFERENCES

Angell, J. K., and J. Korshover, 1983: Global temperature variations in the troposphere and stratosphere: 1958–1982. *Mon. Wea. Rev.*, **111,** 901–921.

Berger, J. O., 1985: *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 617 pp.

Dickey, J. M., 1975: Bayesian alternatives to the *F*-test and least-squares estimates in the normal linear model. S. E. Fienberg and A. Zellner, Eds., *Bayesian Studies in Econometrics and Statistics,* North-Holland.

Ellsaesser, H. W., M. C. MacCracken, J. J. Walton and S. L. Grotch, 1986: Global climatic trends as revealed by the recorded data. *Rev. Geophys.,* **24,** 745–792.

Epstein, E. S., 1982: Detecting climate change. *J. Appl. Meteor.,* **21,** 1172–1182.

——, 1985: *Statistical Inference and Prediction in Climatology: A Bayesian Approach.* American Meteorological Society, Boston.

Hinkley, D. V., 1969: Inference about the intersection in two-phase regression. *Biometrika,* **56,** 495–504.

——, 1971: Inference in two-phase regression. *J. Amer. Stat. Assoc.,* **66,** 736–743.

Jones, P. D., 1985: Southern Hemisphere temperatures: 1851–1985. *Climate Monitor,* **14,** 132–140.

——, S. C. B. Raper and T. M. L. Wigley, 1986: Southern Hemisphere surface air temperature variations: 1851–1984. *J. Climate Appl. Meteor.,* **25,** 1213–1230.

Leamer, E. E., 1978: *Specification Searches.* Wiley and Sons.

Smith, A. F. M., 1975: A Bayesian approach to inference about a changepoint in a sequence of random variables. *Biometrika,* **62,** 407–416.

——, and D. G. Cook, 1980: Switching straight lines: A Bayesian analysis of some renal transplant data. *Appl. Stat.,* **29,** 180–189.

Solow, A. R., 1987: Testing for climate change: An application of the two-phase regression model. *J. Climate Appl. Meteor.,* **26,** 1401–1405.

Wigley, T. M. L., and P. D. Jones, 1981: Detecting carbon dioxide-induced climatic change. *Nature,* **292,** 205–208.