Testing for Climate Change: An Application of the Two-Phase Regression Model

ANDREW R. SOLOW

Woods Hole Oceanographic Institution, Woods Hole, MA 02543

(Manuscript received 1 September 1986, in final form 20 April 1987)

ABSTRACT

A statistical test for detecting a change in the behavior of an annual temperature series is presented. The test is based on the two-phase regression model. By treating the hypothesized time of change as an unknown parameter, the approach allows an inference to be made about the time of change. The approach also avoids a serious problem, called data-dredging, that can arise in testing for change occurring at a specified time. The test is applied to a series of Southern Hemisphere temperatures, and the hypothesis of no change cannot be rejected.

1. Introduction

Epstein (1982), in an interesting and important paper, discussed the problem of detecting climate change in terms of testing statistically for change of a specified form occurring at a specified time. The purpose of the current paper is to describe and apply an alternative statistical test for change that does not require the prior specification of the time of onset. This approach, which is based on the two-phase regression model, has certain advantages over testing for change at a specified time. First, by treating the time of change as an unknown parameter, this approach allows us to make inference about the time of change. Second, from a practical point of view, this approach avoids a serious problem called data-dredging. As Epstein (1982) pointed out, in the context of testing for change occurring at a specified time, data-dredging arises when the hypothesized time of change is chosen by referring to the same data on which the test is carried out. A consequence of data-dredging is that the nominal significance level of the test may be exaggerated substantially. We will return to this point later.

In section 2 we introduce the two-phase regression model, and we describe certain inferential procedures that are of interest in the problem of detecting climate change. In section 3 we apply these procedures to an annual temperature series. We make some concluding remarks in section 4.

2. 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 test for 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 are 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 will, in general, include a discontinuity in the signal at the changepoint. Note that under the general model (1), the prechange climate need not be stationary (i.e., b_0 need not be zero). It is possible to constrain model (1) 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.

Using the two-phase regression model, we will test for climate change. This amounts to testing the null hypothesis H_0: $b_1 - b_0 = 0$ against the two-sided alternative hypothesis H_1: $b_1 - b_0 \neq 0$. If we are interested in testing for change in a particular direction (i.e., warming or cooling), we would test H_0 against the appropriate one-sided alternative. In addition to testing for change, we will make an inference about the changepoint, c. Specifically, we will find the maximum likelihood estimate of c, and we will provide associated confidence intervals.

The two-phase regression model has been studied by Hinkley (1969, 1971). The mathematical development is contained in the first of these papers, while a more general overview is contained in the second. Unfortunately, Hinkley (1971) contains several typographical errors. The approach that we outline herein is equivalent to Hinkley's, although it is formulated differently.

The basis of the approach will be to find the maximum likelihood estimate of c, which we denote $\hat{c}$. Because no closed form expression for $\hat{c}$ is available, it is necessary to search the likelihood function to find its maximum. Model (1) can be rewritten as

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

where

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

and where $b = b_1 - b_0$. For a fixed value of c, (2) is a normal linear regression model with regressor variables i and $(i - c)\,\mathrm{IND}_c(i)$. To find c, we search over possible values to find the one with smallest residual sum of squares. The estimates of a_0, b_0, and b (which we denote $\hat{a}_0$, $\hat{b}_0$, and $\hat{b}$, respectively) are found from fitting model (2) with $c = \hat{c}$. Hinkley (1971) describes a fairly efficient search algorithm, although for small datasets a simple grid search seems reasonable.

To test the hypothesis of no change against the two-sided alternative, we use the likelihood ratio statistic (Hogg and Craig, 1970). For the two-phase regression model, the likelihood ratio statistic is

$$U = [(S_0 - S)/3]/[S/(n-4)] \qquad (3)$$

where S_0 is the residual sum of squares from fitting the null model

$$T_i = a_0 + b_0 i + e_i \qquad (4)$$

and S is the residual sum of squares from fitting the alternative model (2) with $c = \hat{c}$. The asymptotic distribution of U under the null hypothesis of no change is the F distribution with 3 and $n - 4$ degrees of freedom, which we will denote by $F_{3,n-4}$. The test procedure is to reject the null hypothesis at the $1 - \alpha$ level of significance if

$$U \geq F_{3,n-4}(1 - \alpha)$$

where $F_{3,n-4}(1 - \alpha)$ is the $1 - \alpha$ quantile of the $F_{3,n-4}$ distribution.

It is instructive to consider the effect of data-dredging on the performance of this test. If the changepoint is specified independently of the data, the statistic for testing the hypothesis of no change against the two-sided alternative is

$$U' = \frac{S_0 - S}{S/(n-3)} \cong 3U.$$

With c specified, the null distribution of U' is $F_{1,n-3}$, and the test procedure is to reject H_0 at the $1 - \alpha$ significance level if

$$U' \geq F_{1,n-3}(1 - \alpha).$$

Note that if c is estimated by $\hat{c}$ but we act as if c is specified, the test statistic is approximately three times larger than it should be. This effect is partially offset by the fact that the upper quantiles of the $F_{1,n-3}$ distribution are somewhat larger than the corresponding quantiles of the $F_{3,n-4}$ distribution. For example, the ratio of the 0.95 quantiles of the $F_{1,\infty}$ and $F_{3,\infty}$ distributions is approximately 1.48. This has the consequence that if we estimate c by $\hat{c}$, but act as if c is specified and apply the standard F test, we will achieve a level of significance that is spuriously high. For example, suppose that we observe $U' = 4.0$ with large n. If we act as if c is specified we would reject the null hypothesis at the 0.95 level. However, if we act as if c is unknown, the statistic U would have the value 1.33, and we would reject the null hypothesis even at the 0.75 level of significance.

If we wish to test for change in a particular direction (e.g., to test for the onset of warming due to the greenhouse effect), a rough one-sided test can be based on the asymptotic distribution of $\hat{b}$. In model (2), $\hat{b}$ is asymptotically normally distributed with mean b and variance

$$\mathrm{var}\,\hat{b} = \sigma^2(1/C_r + 1/C_{r*})$$

where C_r is the corrected sum of squares for the integers $i = 1, \cdots, r$ and C_{r*} is the corrected sum of squares for the integers $i = r + 1, \cdots, n$. That is:

$$C_r = \sum_{i=1}^{r}(i - m_r)^2, \quad m_r = \sum_{i=1}^{r} i/r$$

and

$$C_{r*} = \sum_{i=r+1}^{n}(i - m_{r*})^2, \quad m_{r*} = \sum_{i=r+1}^{n} i/(n-r).$$

Since σ^2 is unknown, we replace it by its estimate $S/(n - 4)$. The test procedure is to reject H_0: $b = 0$ in favor of H_1: $b > 0$, say, at the $1 - \alpha$ significance level if

$$\hat{b}/[S(1/C_r + 1/C_{r*})/(n-4)]^{1/2} \geq t_{n-4}(1 - \alpha)$$

where $t_{n-4}(1 - \alpha)$ is the $1 - \alpha$ quantile of the Student's t distribution with $n - 4$ degrees of freedom. As Hinkley (1969) pointed out, $\hat{b}$ will be positively biased (in absolute value) in finite samples, because the maximum likelihood procedure is associated with maximizing the difference between the two regression slopes. For fixed n, the importance of this bias is greater for small values of the ratio b/σ. This means that the test just described may be misleading in small samples, since it depends on $\hat{b}$ being unbiased under the null hypothesis H_0: $b = 0$.

Finally, we can use the likelihood ratio statistic U to find a two-sided confidence interval for c. The two-sided test of H_0: $c = c_0$ with asymptotic significance level $1 - \alpha$ is to accept H_0 if

$$(S' - S)/[S/(n-4)] \leq F_{1,n-4}(1 - \alpha) \qquad (5)$$

where S' is the residual sum of squares from fitting model (2) conditional on $c = c_0$. The $(1 - \alpha)$ confidence interval for c is the set of c_0 satisfying (5). We can take

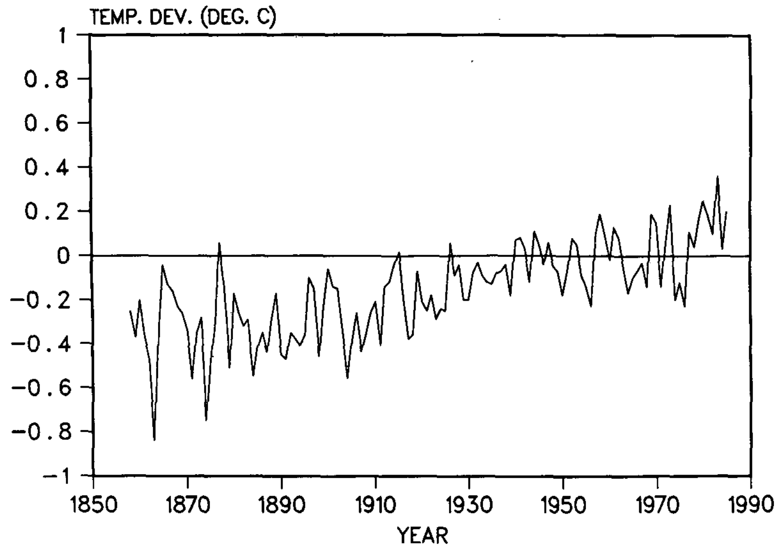

FIG. 1. Plot of annual Southern Hemisphere temperature deviations, 1858–1985 (Jones, 1985).

the endpoints of the confidence interval to be the roots of

$$S = S' - SF_{1,n-4}(1-\alpha)/(n-4).$$

These roots can be found by the same grid search used to find $\hat{c}$.

3. Application

In this section, we apply the inferential procedures described in section 2 to an annual series of Southern Hemisphere temperature deviations compiled by Jones (1985). 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 was constructed and problems relating to its quality were discussed by Jones et al. (1986).

The residual sum of squares function for fitting model (2) is shown in Fig. 2. The maximum likelihood estimate of c corresponds to the year 1887. The likelihood ratio statistic U has the value 1.88, with 3 and 124 degrees of freedom, which is significant only at the 0.85 level or lower. Based on standard notions of significance, we would not reject the null hypothesis of

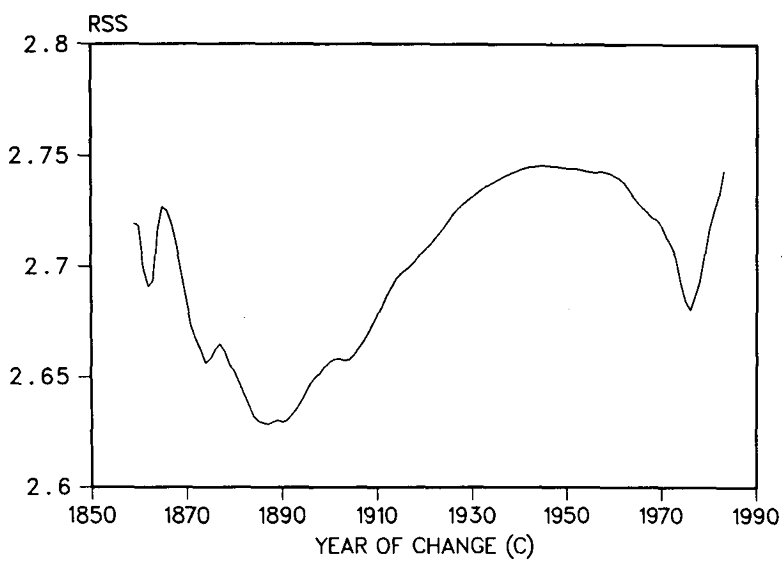

FIG. 2. Residual sum of squares function from fitting model (2) with the changepoint at c.

no change. To illustrate the effect of data-dredging, if we test for change beginning after 1887, but act as if the changepoint was specified independent of the data, the value of the statistic U' is 5.70, with 1 and 125 degrees of freedom. This result is significant at the 0.98 level or lower. Based on standard notions of significance, we would reject the null hypothesis of no change. Clearly, data-dredging has a serious impact on inference in this case.

It is interesting to note that the residual sum of squares function in Fig. 2 has three local minima—corresponding to the years 1862, 1874 and 1976—in addition to the global minimum in 1887. This kind of behavior is common in changepoint problems. Referring to Fig. 1, it is clear that the local minima in 1862 and 1874 are associated with highly local effects. A somewhat weaker local effect also occurred around 1976. Note that the global minimum at 1887 is not associated with any clear local event.

For the two-phase regression model (2), the estimates of b_0 and b are -0.00076 and $0.00554°C\ yr^{-1}$, respectively. The standard error of $\hat{b}$ is 0.00311°C. The t statistic for testing for change has the value 1.78, with 124 degrees of freedom, which is significant at the 0.91 level or lower. Again, based on standard notions of statistical significance, we would not reject the null hypothesis of no change. For the null model (4), the estimate of b_0 is $0.00403°C\ yr^{-1}$, with standard error 0.000353°C. The fitted model explains 49% of the variation in temperature over the observation period.

It is interesting to note that the result based on the asymptotic distribution of b is more significant than that based on the likelihood ratio statistic U. This is explained in the following way. The results of the likelihood ratio test suggest that the null hypothesis is true. As we have said, b is positively biased (in absolute value) for finite samples when the null hypothesis is true. This positive bias causes the significance of the t statistic to be exaggerated. For situations in which the likelihood ratio statistic is more highly significant (or the time series is longer), this effect would be smaller, and the two tests would give similar results.

Turning to inference about the changepoint, an approximate 0.95 confidence interval for c corresponds to the interval 1867 to 1923. This interval is not centered on $c = 1887$. This is due to the fact that the residual sum of squares function rises somewhat faster before 1887 than after. The great width of this interval further reflects the lack of strong evidence for change in this dataset.

4. Discussion

Generally speaking, in order for a statistical procedure to perform well, the assumptions underlying the procedure must be met. For the two-phase regression model, the underlying assumptions are that the model is correctly specified [i.e., that (1) holds], and that the noise terms are independent and normally distributed with constant variance. Although these assumptions may be violated in a number of ways, we will consider only three special cases here.

It is almost certainly true that climate change would not occur abruptly, as in the two-phase regression model, but would include a period of gradual transition. Provided that the period of transition is short relative to the length of the record, the two-phase regression model will perform well. For a record with a relatively long transition period, the two-phase regression model will tend to identify a changepoint that occurs somewhere in the middle, rather than at the beginning, of the transition period. In particular, if the transition period begins at the very end of the record, the two-phase regression model may fail to identify a changepoint altogether. Bacon and Watts (1971) considered the problem of gradual transition in the two-phase regression model.

In addition to assuming an abrupt change, the two-phase regression model assumes, at most, a single change. It is possible in principle to extend the two-phase regression model to allow for two or more changepoints. For example, the maximum likelihood estimates of the two changepoints in a three-phase regression model would be found by a search procedure similar to that used in the two-phase model. For the data shown in Fig. 1, if we require that the two changepoints in a three-phase regression model be at least 10 yr apart, the maximum likelihood estimates of the two changepoints correspond to the years 1888 and 1978, with a residual sum of squares 2.59 $(°C)^2$. In general, the maximum likelihood estimates of the two changepoints in a three-phase regression model need not correspond to local minima in the residual sum of squares function from fitting a two-phase model, although this is the case with the present dataset. The requirement that the two changepoints be separated by several years is needed to reduce the influence of highly local events. For example, if we relax this restriction, the maximum likelihood estimates of the two changepoints correspond to the years 1863 and 1864, and are clearly associated with a local event. Although it may seem reasonable to suppose that an F test could be devised to test the three-phase model against the two-phase model, this is not at all clear. The mathematical analysis of changepoint problems has proved to be extremely difficult, and in some cases intractable (Hinkley, 1969). Indeed, although the procedures described in this paper for making inference under the two-phase regression model are straightforward, the distributional results are based on empirical study, rather than mathematical analysis (Hinkley, 1969). Beyond mathematical considerations, introducing a model with multiple changepoints also raises the possibility of data-dredging in the specification of the number of changepoints in the model.

The autocorrelation function of the residuals from the fitted null model (4) is presented in Fig. 3. Inci-

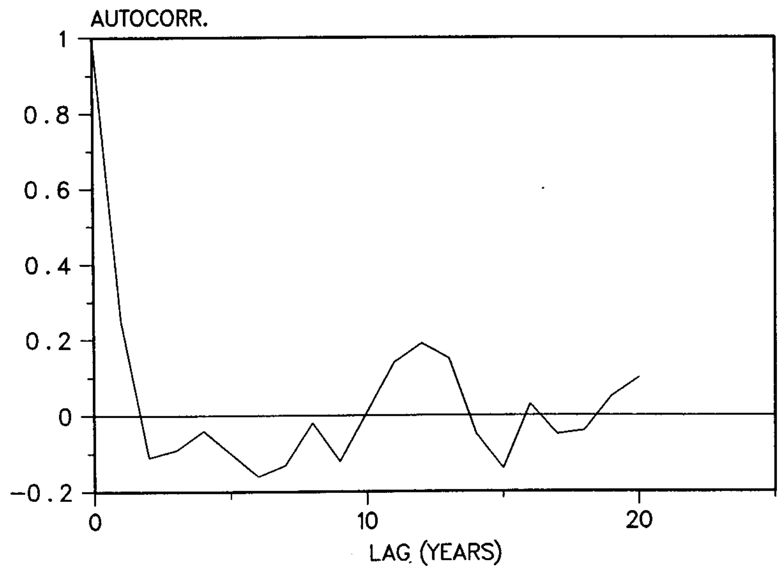

FIG. 3. Autocorrelation function for the residuals from fitting model (4) to the data shown in Fig. 1.

dentally, the autocorrelation function of the residuals from the fitted alternative model (2) is quite close to that shown in Fig. 3. This figure indicates the possible presence of weakly periodic behavior in the residuals, in addition to the possibility of behavior consistent with low-order, ARMA-type structure. While a complete treatment of this problem is beyond the scope of this paper, we make the following remarks about the effects of periodicity on the performance of the two-phase regression model. The sensitivity of the two-phase regression model to periodicity in the noise process depends on both the period and the amplitude of the periodic component. If the period is short relative to the record length, the two-phase model will still perform well. However, if the period is relatively long, a turning point in the periodic component may be identified as a changepoint in the two-phase model. Alternatively, a true changepoint that coincides with a turning point in the opposite direction in the noise will tend to be masked. If the amplitude of the periodic component is small relative to the signal, the two-phase model will work well. However, if the amplitude is relatively large, important end-effects can occur. For example, if the record ends at the top or bottom of a cycle in the noise process, the two-phase model may identify a changepoint late in the record where none exists. Alternatively, a true changepoint late in the record may be masked if the record ends at the top or bottom of a cycle.

An ad hoc procedure for dealing with periodicity would consist of forming the residuals from model (2), fitting a periodic curve to these residuals, subtracting the fitted values from the original data, and applying the two-phase model to the corrected data. This approach is similar to that taken by Rust and Kirk (1978) in their analysis of a time series of atmospheric carbon dioxide concentrations, although they were not concerned with detecting a changepoint. This approach requires that the residuals from the initial fit of model (2) be representative of the true noise process. This will not be the case if the periodicity in the noise process strongly influences the initial fit of model (2).

Acknowledgments. Research was supported by the J. N. Pew, Jr. Charitable Trust and the Marine Policy Center of the Woods Hole Oceanographic Institution. The author would like to thank Dr. James M. Broadus, Dr. Richard W. Katz, and Dr. Edward S. Epstein for their help, and an anonymous reviewer for valuable comments. Woods Hole Oceanographic Institution Contribution Number 6278.

REFERENCES

Bacon, D. W., and D. G. Watts, 1971: Estimating the transition between two intersecting straight lines. *Biometrika,* **58,** 525–534.

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

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.

Hogg, R. V., and A. T. Craig, 1970: *Introduction to Mathematical Statistics.* third ed., Macmillan, 415 pp.

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.

Rust, B. W., and B. L. Kirk, 1978: Inductive modelling of population time series. *Time Series and Ecological Processes,* H. H. Shugart, Jr., Ed., SIAM, 154–192.