Discriminating between Models: An Application to Relative Sea Level at Brest*

ANDREW R. SOLOW

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Manuscript received 26 July 1986, in final form 18 October 1989)

ABSTRACT

A simple, informal method is presented for discriminating between competing models of trend in a climate record. The method is applied to a tide gauge record of relative sea level at Brest for the period 1807–1970. Although relative sea level at Brest appears to have accelerated over this period, it is impossible to distinguish between a smooth acceleration and one that occurred over a relatively short period of time.

1. Introduction

Many long-term records of climatological variables exhibit trends. It is common to estimate such trends by regressing the variable of interest against specified functions of time. This kind of analysis sometimes requires choosing between competing functional forms of the trend. In some cases, the competing models are nested in the sense that they coincide for some choice of parameter values. For example, the linear trend:

$$E(Y_t) = \alpha_0 + \alpha_1 t \tag{1}$$

and the quadratic trend:

$$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 t^2 \tag{2}$$

are nested because they coincide if $\beta_2 = 0$. Here, Y_t is the value of the variable of interest in period t and E is the expectation operator. A well-developed theory exists for choosing between nested models (e.g., Silvey 1970). For example, under the usual normal theory assumptions about the errors in Y_t, a standard F test (or, equivalently, a t test) can be used to choose between (1) and (2) by testing the null hypothesis $\beta_2 = 0$.

In some cases, the competing models are not nested. For example, the two-phase linear regression model has been used in the climatological literature to represent a change in trend (e.g., Solow 1987; Hanson et al. 1989). This model can be written

$$E(Y_t) = \Theta_0 + \Theta_1 t + \Theta_2(t - \Theta_3)I(t; \Theta_3) \tag{3}$$

where Θ_3 is an unknown change-point and

$$I(t; \Theta_3) = \begin{cases} 0, & \text{if} \quad t \leqslant \Theta_3 \\ 1, & \text{otherwise.} \end{cases}$$

The fundamental difference between the two-phase model and the quadratic model is that the former exhibits a distinct change from one regime to another, while the latter does not. While it is more likely from a physical point of view that a period of transition exists between regimes, the two-phase model is still useful provided that the transition period is short.

Note that, although the linear model is nested within both the quadratic model and the two-phase model, the quadratic model and the two-phase model are themselves nonnested. For this reason, it is not possible to choose between these two models by testing a simple null hypothesis. A related difficulty is that, unlike the case of nested models, there is no natural choice of the null model.

Cox (1962) and Atkinson (1970) considered the problem of choosing between nonnested models. Cox (1962) distinguished between hypothesis testing and discrimination. In hypothesis testing, the reasonableness of the null model is assessed in light of a specific alternative. In discrimination, the competing models are on an equal footing. In the discussion of Atkinson (1970), Williams (1970) outlined an approach to discriminating between models based on simulation. The purpose of this paper is to extend this approach slightly and to apply it to the analysis of an annual tide gauge record of relative sea level at Brest.

Alternative methods for choosing between models include Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). These are discussed by Katz (1981) in the context of choosing the order of a Markov chain model for meteorological data.

* Woods Hole Oceanographic Contribution Number 7233.

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

2. Approach

Suppose that the two competing models are

$$Y_t = g(t; \alpha) + e_t \quad (4)$$

and

$$Y_t = h(t; \beta) + f_t \quad (5)$$

where α and β are vectors of unknown parameters and the e_t's and f_t's are independent normal errors with mean zero and unknown variances σ_e^2 and σ_f^2, respectively. The case of nonnormal data is discussed briefly in section 4.

Williams (1970) suggested the following procedure for discriminating between (4) and (5) from observations for $t = 1, \cdots, n$. Estimate the parameters α and β by fitting each model to the data. Form the ratio of residual sums of squares:

$$r = \mathrm{RSS}_g/\mathrm{RSS}_h = \sum (Y_t - g(t; \alpha^*))^2 / \sum (Y_t - h(t; \beta^*))^2 \quad (6)$$

where α^* and β^* are the estimates of α and β. This ratio will be used for discriminating between the two models.

The discrimination procedure essentially addresses the question: Is the observed value of r consistent with sampling from (4) or from (5)? In many cases, it is also possible that neither (4) nor (5) is the correct model, or that it is impossible to distinguish between (4) and (5) on the basis of the data in hand. The discrimination method should also allow for these conclusions.

In order to discriminate between (4) and (5) using r, it is necessary to know how different r can be under different realizations of the two models. That is, it is necessary to know the sampling distribution of r under both (4) and (5). These distributions can be estimated by simulation. For example, to estimate the sampling distribution of r under (4), simulated values are generated according to

$$Y_t^s = g(t; \alpha^*) + e_t^s \quad t = 1, \cdots, n$$

where e_t^s is a value drawn at random from the normal distribution with mean zero and variance:

$$s_e^2 = \mathrm{RSS}_g/(n - p_\alpha) \quad (7)$$

where p_α is the number of elements of α. A value of r can be found by fitting (4) and (5) to the simulated record. The entire procedure can be repeated a large number of times and the sampling distribution of r under (4) can be estimated from the values of r generated in this way. The same procedure can be used to estimate the sampling distribution of r under (5).

Once these sampling distributions have been estimated, it is straightforward to draw one of four conclusions: the observed value of r is consistent with sampling from (4), from (5), from neither (4) nor (5), or from both (4) and (5).

The procedure outlined above discriminates between two models using a single measure of the relative goodness-of-fit of the two models. A potential problem arises if both absolute fits are poor. In that case, the measure of the relative fit can be arbitrary. To avoid this problem, the procedure can be extended to discriminate between the models using the pair of residual sums of squares (RSS_g, RSS_h). In this case, discrimination is based on the estimated bivariate sampling distribution of the residual sums of squares under the two models. This extension also allows, in principle, the generalization of the procedure to discriminating among more than two models. For example, discrimination among three models can be based on comparing the triplet of residual sums of squares to their sampling distributions under each of the three models.

The discrimination procedure can be applied in a formal manner, by using a loss function and estimated probabilities of different discrimination errors. In this paper, however, the approach will be used as an informal, graphical guide, rather than as a formal decision rule. In the case of discriminating among several models, however, a graphical procedure is precluded and some kind of formal or semiformal rule must be applied.

3. Application

Figure 1 shows an annual tide gauge record of relative sea level in Brest, France. The period of observation runs from 1807 to 1970, and there are 23 missing observations. It is clear from Fig. 1 that relative sea level has been rising in Brest.

For the purposes of illustration, consider choosing between the linear trend model (1) and the quadratic trend model (2). Because these models are nested, it is unnecessary to use the method outlined in section 2. The two models were fit to the data in Fig. 1 and (with t scaled to run from −1.0 in 1807 to 1.0 in 1970) the parameter estimates were

$$\alpha_0^* = 6964.8 \quad \alpha_1^* = 73.9 \quad \mathrm{RSS}_g = 198494.6 \quad s_e = 37.8$$

$$\beta_0 = 6949.4 \quad \beta_1^* = 76.9 \quad \beta_2^* = 47.3 \quad \mathrm{RSS}_h = 167174.3$$

$$s_f = 34.8.$$

The fitted models are shown in Fig. 1.

A total of 100 records were simulated from each of the two fitted models. For each simulated record, RSS_g (corresponding to the linear model) and RSS_h (corresponding to the quadratic model) were found. These are plotted in Fig. 2. When simulating from the linear model, RSS_g and RSS_h tend to be close. On the other hand, when simulating from the quadratic model, RSS_g

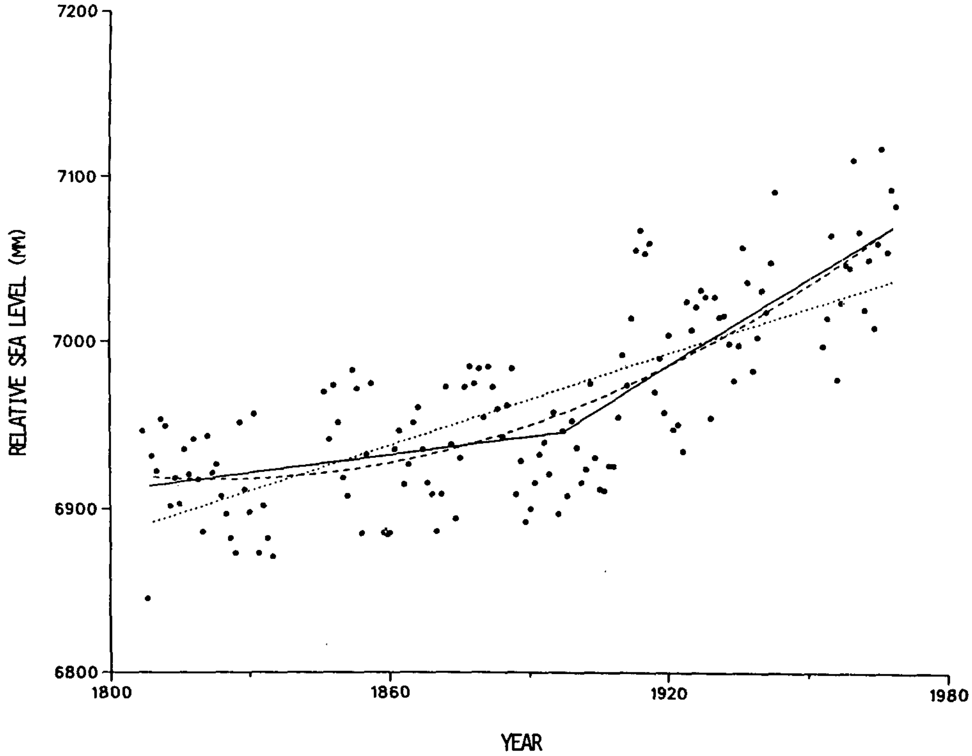

FIG. 1. Relative sea level at Brest, 1807–1970. Also shown are the estimated linear (dotted), quadratic (dashed), and two-phase (solid) trend models.

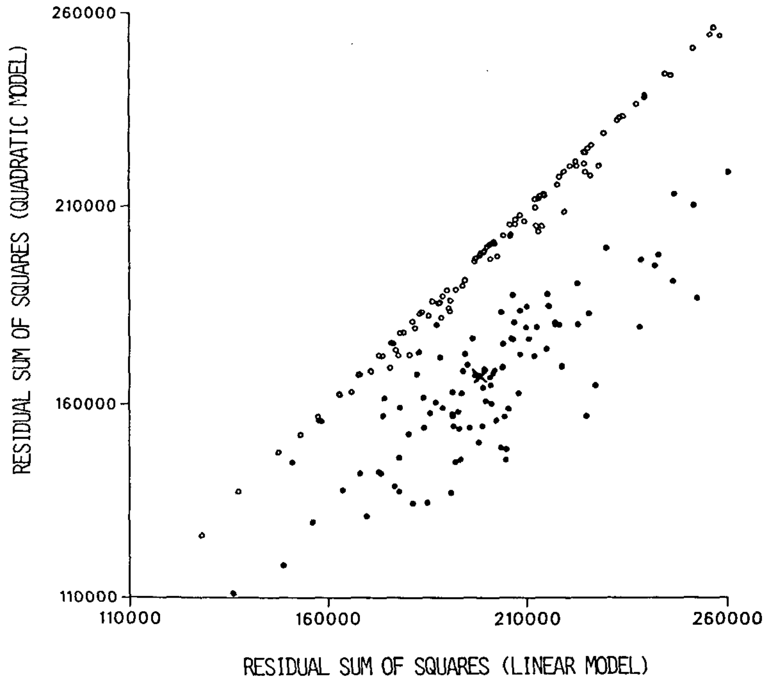

FIG. 2. Residual sums of squares for the quadratic model versus residual sums of squares for the linear model for realizations from the linear model (open circles) and the quadratic model (closed circles). Also shown is the observed value of the residual sums of squares (cross).

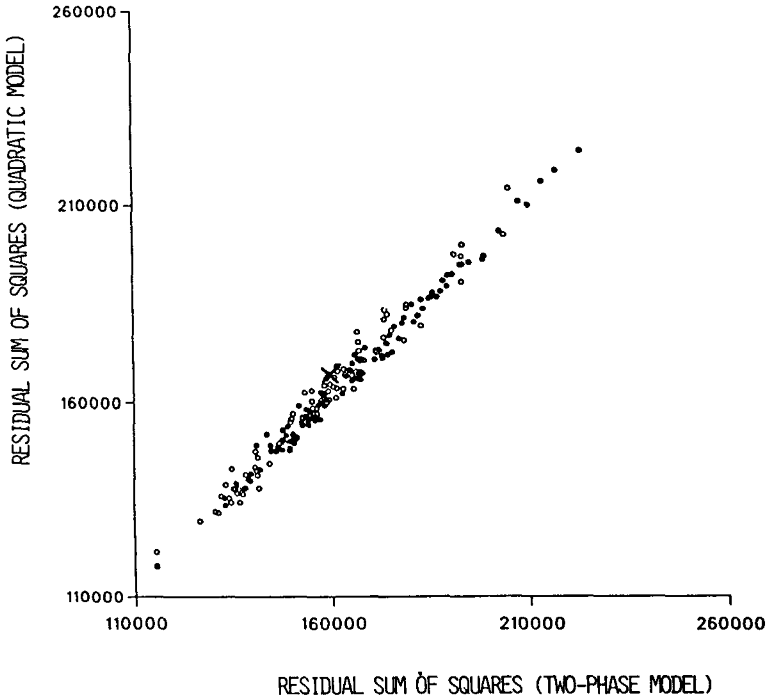

FIG. 3. Residual sums of squares for the quadratic model versus residual sums of squares for the two-phase model for realizations from the two-phase model (open circles) and the quadratic model (closed circles). Also shown is the observed value of the residual sums of squares (cross).

tends to be substantially higher than RSS_h. The observed pair of residual sums of squares, which is also shown in Fig. 2, is clearly inconsistent with sampling from the linear model and clearly consistent with sampling from the quadratic model. This conclusion is borne out by a standard test of the significance of β_2^*.

Next, consider choosing between the quadratic trend model (2) and the two-phase linear regression model (3). The two-phase model was fit to the data in Fig. 1 and the parameter estimates were

$$\Theta_0^* = 6943.2,\ \Theta_1^* = 29.2,\ \Theta_2^* = 111.1,\ \Theta_3^* = 1897$$

$$RSS_g = 159803.3,\ s_e = 34.2.$$

The fitted model is also shown in Fig. 1.

The simulation procedure was applied as before, and the results are shown in Fig. 3. These results indicate that the bivariate sampling distributions of (RSS_g, RSS_h) under the two fitted models are very close. The observed value of (RSS_g, RSS_h) is also shown in Fig. 3 and appears to be slightly more consistent with the two-phase model, although it is impossible to distinguish clearly between the two models on the basis of these data. This result is not surprising given the closeness of the fits for the two models.

4. Discussion

This paper has presented a simple, informal method for discriminating between competing models. Although the discussion centered on the problem of trend estimation, the method can be applied to a wide variety of problems involving model choice. For example, Solow (1989) discussed the use of this approach in the comparison of climate model output and climate data. A different kind of application would be discriminating between two competing models of the distribution of a variable.

As described in section 2 and as applied in section 3, the discrimination method assumed that the data are normally distributed. It is straightforward to modify the method to handle nonnormal data. In particular, two modifications are needed. First, discrimination should be based on a statistic suited to the distribution of the data. A good choice of statistic is the log-likelihood (Silvey 1970) which, in the case of normal data, corresponds to the residual sum of squares. Second, in the simulation procedure, the observations should be simulated from the assumed distribution and not from a normal distribution. Similar modifications can be made if the data exhibit serial correlation or nonconstant variance.

In the application to relative sea level at Brest, the method discriminated clearly in favor of the quadratic model over the linear model. However, it was impossible, to discriminate between the quadratic model and the two-phase model. From a practical point of view, this means that it is unclear whether the apparent acceleration in relative sea level at Brest occurred over a relatively short period of time near the turn of the century or whether it occurred smoothly over the course of the data.

Acknowledgments. I thank two anonymous reviewers for their helpful comments. This work was supported by the J.N. Pew Charitable Trusts and by the National Sea Grant College Program under Grant NA86-AA-D-SG090 (R/0-4).

REFERENCES

Atkinson, A. C., 1970: A method for discriminating between models. *J. Roy. Stat. Soc.*, **B32,** 323–353.

Cox, D. R., 1962: Further results on tests of separate families of hypotheses. *J. Roy. Stat. Soc.*, **B24,** 406–424.

Hanson, K., G. Maul and T. Karl, 1989: Are atmospheric "greenhouse" effects apparent in the climatic record of the contiguous United States (1895–1987)? *Geophys. Res. Lett.*, **16,** 49–52.

Katz, R. W., 1981: On some criteria for estimating the order of a Markov chain. *Technometrics,* **23,** 243–249.

Silvey, S. D., 1970: *Statistical Inference.* Chapman and Hall, 191 pp.

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

——, 1989: On the statistical comparison of climate model output and climate data. *Proc. of the Department of Energy Workshop on Greenhouse Gas Induced Climate Change,* in press.

Williams, D. A., 1970: Discussion of Dr. Atkinson's paper. *J. Roy. Stat. Soc.*, **B34,** 350.