Program Abstracts/Algorithms

A program for fitting two-phase segmented-curve models with an unknown change point, with an application to the analysis of strategy shifts in a cognitive task

A. LEO BEEM

Leiden University, Leiden, The Netherlands

A program is described for fitting a regression model in which the relationship between the dependent and the independent variables is described by two regression equations, one for each of two mutually exclusive ranges of the independent variable. The point at which the change from one equation to the other occurs is often unknown, and thus must be estimated. In cognitive psychology, such models are relevant for studying the phenomenon of strategy shifts. The program uses a (weighted) least squares algorithm to estimate the regression parameters and the change point. The algorithm always finds the global minimum of the error sum of squares. The model is applied to data from a mental-rotation experiment. The program's estimates of the point at which the strategy shift occurs are compared with estimates obtained from a nonlinear least squares minimization procedure in SPSSX.

The last decade has seen substantial interest in the application of cognitive-process models of task performance to the analysis and description of individual differences in performance (see Lohman & Ippel, 1993; Snow & Lohman, 1993). One of the determinants of task performance that has emerged from these investigations is the strategy-shift phenomenon, whereby a subject may apply different strategies for solving different items of the same task. Studies that have identified strategies employed by subjects have used a variety of measures, including retrospective reports, eye-fixation records, errors, and measures of response latencies (e.g., Cooper, 1980; Glushko & Cooper, 1978; Kyllonen, Lohman, & Snow, 1984; Snow, 1980).

Kyllonen, Lohman, and Woltz (1984) systematically studied subjects' strategy shifts on different items of the same spatial task as a function of item characteristics. They fitted several regression models for all subjects in order to identify the strategies that subjects used and the possible strategy shifts that occurred. Each model postulated different processes underlying performance, and some models postulated strategy shifts (of different

kinds). The strategy-shift models were fitted with a few different shift points that were fixed a priori and that were the same for all subjects, rather than being estimated from each subject's data separately. The best-fitting process model for each subject was determined by comparing adjusted R-square statistics from different models. The results suggested not only that different subjects use different strategies, but also that the same subjects may use different strategies on different items of the same task, as well as that subjects differ in the point at which they shift from one strategy to another.

Ippel and Beem (1987) fitted latencies of subjects on a mental-rotation task. In mental-rotation tasks, subjects must determine whether two figures are the same in shape or are mirror images. The figures can differ in orientation. Latencies typically suggest that subjects mentally rotate one of the figures into congruence with the other. On each trial, rotation can be achieved by either a clockwise or a counterclockwise routine. Depending on the angular disparity of the figures, one of these routines is objectively the most efficient, whereas 180° of angular disparity is an objectively optimal point for shifting from one rotation routine to the other. Ippel and Beem fitted subjects' latencies as a linear function of angular disparity with a model for which the shift point was not assumed to be known a priori, but was estimated from each subject's data separately. Latencies' linear dependence on angular disparity implies that speed of rotation is constant for a given rotation routine, although the rotation speeds of the two routines may differ in the model. The subjects appeared to differ substantially with respect to the angular disparity at which a strategy shift occurred. Performance indices derived from the model parameters were correlated with other variables. It was found that the absolute value of the difference between the subjects' estimated shift points and the objectively optimal shift point correlated with measures of non-verbal intelligence.

The model that Ippel and Beem (1987) used belongs to a class of models variously called segmented-curve models, piecewise regression models, or switching regression models. In the present article, a computer program, Segcurve, will be described that can be used to fit such segmented-curve models under the assumption that the strategy-shift point is unknown for each subject, and thus must be estimated. Segmented-curve models arise quite frequently in many domains. In general, they are appropriate whenever the relationship between the dependent and independent variables changes as a function of some (other) variable. For example, Woods, Resnick, and Groen (1977) suggested that children use one of two mental-subtraction strategies to solve the problem

Requests for reprints should be sent to A. L. Beem, Centre for the Study of Education and Instruction, Leiden University, Wassenaarseweg 52, P.O. Box 9555, 2300 RB Leiden, The Netherlands (e-mail: beem@rulfsw.leidenuniv.nl).

$m - n =$; that is, they either count up from the smaller number or count down from the larger number. Children appeared to be inclined to apply the strategy that is (approximately) most efficient for the problem at hand (i.e., whether they count $m - n$ times up from n to m or count n times down from m depends on which strategy requires fewer counts). Response latencies as a function of $m - n$ should then be fitted with a two-phase segmented-curve model. The models may also be useful to study the course of learning. Task performance or the relationship between performance and other variables is often observed to change (qualitatively) as a function of time. The time scale may span several years, as in developmental studies, or it may be relatively short. Mathews, Buss, Chinn, and Stanley (1988) fitted segmented-curve models to performance measures in a concept-learning task as a function of trial-block number, in order to estimate in which trial block individual subjects' performance would change from chance level to a transition phase, and from the latter to an asymptotic level. (Since I have not been able to find the published version of the submitted manuscript to which Mathews et al. refer, and which describes their program and the algorithm used, I cannot compare their program with Segcurve.)

In this paper, I will first discuss the model, the conditions under which it may be applied, and some extensions. After that, I will describe in general terms the program's estimation method, some of its options, and its input and output; the program manual (Beem, 1993) contains detailed information. Finally, the model will be applied to data from a mental-rotation experiment, and the estimates of the shift points compared with estimates obtained from a procedure for nonlinear regression in SPSSX.

TWO-PHASE SEGMENTED-CURVE MODELS

The following notation is used in describing the model: y = dependent variable; x = independent variable; m = number of levels of x (i.e., the number of different values of x); n_i = number of observations at Level i of x ; y_{ij} = j th observation on y at Level i of x ; and x_{ij} = j th observation on x at Level i of x .

If the model is applied to data from a mental-rotation experiment, y would be the response latency and x the angular disparity between the presented figures. It is assumed that

$$x_{ij} > x_{kl} \text{ iff } i > k, \text{ and } x_{ij} = x_{kl} \text{ iff } i = k,$$

where $i, k = 1, \dots, m, j = 1, \dots, n_i, l = 1, \dots, n_k$. If only the level of x , or only the level of x at which y is observed, is relevant, subscripts j will be dropped.

The program assumes that subjects use two different strategies for a given task. One strategy is used if $x \leq s$; the other is used if $x > s$. Performance can then be described by a two-phase model, in which the relationship between dependent and independent variables is described by two regression equations, one for each of two

mutually exclusive ranges of the independent variable, as in Equations 1b and 1c below.

The number of strategies needed to describe subjects' task behavior depends, at least partly, on the level at which process components are described, and hence also on the way in which the (experimentally manipulated) task facets are related and arranged. For many tasks, the level of description will be such that a two-phase model is a priori the most plausible alternative to a one-strategy model. However, segmented-curve models can, in principle, be formulated for any number of strategies. Because the program does its computations on matrices and vectors, it is fairly easily extended to more than two strategies. Such extensions basically involve the addition of an inner loop to the program. For various reasons, it seems more appropriate to implement other models in different versions of the program as the need arises rather than to develop a program that can handle all conceivable special cases (one such reason being that the number of different join and jump models, which are described below, rapidly grows with an increasing number of strategies).

A psychological-process model often postulates a processing phase (e.g., an encoding phase) whose execution time is assumed to be constant from trial to trial. Some tasks allow for the direct estimation of the execution time for such a phase. In that case, a special hypothesis becomes of interest, and will be explained in the context of mental rotation. In mental-rotation experiments, x values of 0 and 360 (degrees of angular disparity) both correspond to orientation angles of 0°, but they differ conceptually as the values of x at which the lower bounds of response times under the two rotation strategies (clockwise or counterclockwise) are attained. It may be of interest to test not only whether these lower bounds are equal to each other, but also whether they are equal to the response time when subjects do not need to rotate but only to encode the figures (i.e., when the angular disparity is 0°, which means that $x = 0$). To allow for an investigation of the latter hypothesis, the program can fit a constant when observations are available at $x = 0$. Therefore, below, 1a is added to the two-phase model proper (i.e., 1b and 1c). Thus, in its most general form, the program fits the following model:

$$y_{1j} = a_0 + e_{1j} \quad (1a)$$

$$y_{ij} = a_1 + b_1 x_{ij} + e_{ij}, x_1 < x_i \leq s \quad (1b)$$

$$y_{ij} = a_2 + b_2 x_{ij} + e_{ij}, s < x_i \leq x_m, \quad (1c)$$

where, in 1a, y_{1j} is the j th observation at $x_1 = 0$, and a_0 denotes the mean response time for $x = 0$, in which case subjects do not need to rotate, by hypothesis. 1b corresponds to the first strategy, which is employed up to $x = s$, the point at which the regression model changes from 1b to 1c (from here on, the more general term *change point* will be used instead of shift point). In 1b, the parameter a_1 denotes the mean response time for all processes, other than rotation, that are constant over

trials, and b_1 denotes the inverse of the speed of rotation for a counterclockwise rotation. In 1c, which corresponds to the second strategy, $a_2 + b_2 * 360$ is the mean response time for all processes, other than rotation, that are constant over trials, and the absolute value of b_2 is the inverse of the rotation speed for a clockwise routine. The random variables e are the disturbances or error terms, which, as usual, have zero expectation.

Within the class of two-phase models represented by 1a–1c, several distinctions can be made, the first of which is that they may or may not be continuous in the point s . If the model is continuous, it is called a *join* model, because the value of y at s is the same for 1b as it is for 1c. If it is not continuous in s , the model is called a *jump* model. Figures 1 and 2 are examples, respectively, of a join model and a jump model. The values of x at which observations are supposed to have been obtained are shown on the x -axis. For both the join and the jump model, the highest *observed* value of x in 1b is $x = 200$, and its lowest *observed* value in 1c is $x = 240$. For the join model, the change point s occurs in the interval between $x = 200$ and $x = 240$, at $x = 220$, which is not a value at which observations are obtained. The lower and upper bounds of the interval in which the change occurs are denoted, respectively, by x_L and x_U . Thus, $x_L = 200$ and $x_U = 240$. But since the model is assumed to be continuous at the change point, the change point can be estimated as the point at which the regression 1b and 1c cross. However, the change point of the jump model can

Figure 1. Example of a join model with change point at 220° of angular disparity.

Figure 2. Example of a jump model with change point at 200° of angular disparity.

lie anywhere in the closed interval from x_L to x_U , except at $x = 220$, because the error sum of squares remains the same, regardless of the exact location of the change point. Therefore, s is not precisely defined for a jump model. The point $x = 200$ is regarded here as the change point (i.e., the point after which the change from 1b to 1c occurs).

In both the join model and the jump model, the regression lines 1b and 1c cross at some point. This point is called the *meet* in the sequel and in the program's output. In Figures 1 and 2, the meet occurs, respectively, at $x = 220$ and $x = 160$. In the join model, but not in the jump model, the meet coincides with the change point.

The distinction between the join and jump models is not merely a formal one—it may also have psychological implications. Ippel and Beem (1987) proposed that subjects rotate counterclockwise up to the point at which clockwise rotation requires less rotation time. Efficient performance then requires a strategy shift. This hypothesis implies a join model. If a jump model fits better, subjects do not switch strategies optimally, given their speed of clockwise and counterclockwise rotation (assuming rotation speed is not dependent on the rotation angle).

A second distinction between the models is based on a characteristic of the independent variable. For mental-rotation data, the independent variable is essentially continuous. However, it may also be essentially discrete. For example, a discrete segmented-curve model may be ap-

plied to the magnitude comparison of two-digit numbers, where a change of the first digit may or may not change the function relating numbers to their subjective magnitude (cf. Hinrichs, Yurko, & Hu, 1981). In counting models for mental arithmetic, the independent variable (i.e., the number of counts) is also essentially discrete, as in the models of Woods et al. (1977). Thus, in addition to the distinction between join models and jump models, which refers to the continuity or discontinuity of the model in the change point, there is also a distinction made on the basis of whether a model has a continuous or a discrete independent variable.

In a join model with a continuous independent variable, the change point can occur anywhere in the closed interval from x_L to x_U (i.e., $x_L \leq \text{meet} \leq x_U$), and thus also at x_L or x_U (actually, the solution with minimum error sum of squares is sometimes obtained by restricting the change point to occur at one of the observed values of x ; when this happens, it is reported by the program). In contrast, the change point of a model with a discrete independent variable should occur only at the restricted set of values that are prescribed by the theoretical model. In the previous examples, the set consists of a subset of the positive whole numbers (e.g., the number of counts). The program can be instructed to find such discrete solutions, but only for those values that are actually present in the data.

A jump model with a continuous independent variable is conceptualized differently from a jump model with a discrete independent variable. Suppose that, for the model with a continuous independent variable, the change from 1b to 1c occurs for values of x greater than some point x_i (i.e., $x_i = x_L$, the highest observed value in 1b). This implies that the meet of the join model lies somewhere in the closed interval from x_i to x_{i+1} . The same model can be regarded as a jump model if the regression lines 1b and 1c are assumed not to enter the interval from x_i to x_{i+1} . However, this is an artificial model for an essentially continuous variable. The meet of a jump model for a continuous independent variable is therefore assumed to occur outside the interval from x_j to x_{j+1} , $j \neq i$ (note that x_j is by definition different from the corresponding point x_i of the join model). In a discrete model, it is not artificial that the meet can occur in the interval from x_L to x_U , since the regression is not defined within the interval. For example, the model in Figure 1 can be interpreted as a jump model if the independent variable is discrete. Thus, the meet for a discrete jump model may lie either inside or outside the interval from x_L to x_U . However, if the meet lies inside the interval, it should, of course, not be interpreted in the same way as the corresponding change point of the join model for a continuous independent variable. In Figure 1, the open interval from $x = 200$ to $x = 240$ would not exist for the discrete jump model. The program manual (Beem, 1993) discusses the consequences of the continuity or discreteness of the independent variable for the relative fit of the jump model and the join model.

CONDITIONS FOR APPLICABILITY AND POTENTIAL EXTENSIONS OF THE MODEL

To apply the model, it must be possible to order the levels of the independent variable on an interval scale according to some relevant characteristic (which is often some objective definition of the difficulty or complexity of an item). Some measure of processing effort, such as error frequencies or latencies, should be available as a dependent variable. When the dependent variable is a proportion (e.g., proportion correct), the arcsine transformation or the logarithm of the odds ratio, $\log p / (1 - p)$, may be more appropriate.

In one of the strategy models fitted by Kyllonen et al. (1984), the strategy shift consisted of employing a verbal rather than a spatial strategy (or vice versa) to solve a problem. A difficulty measure of a problem was available for each of the strategies. Although at present the program cannot fit the model to such data, the code can be easily adapted to this purpose. The change that needs to be made relates to the ordering of the data, which would have to be done on the basis of the difference between the verbal and spatial difficulty measures, while keeping the difficulty measures as the independent variables.

PARAMETER ESTIMATION

Hudson (1966) describes the estimation method that is used in the program. This is a least squares method, which always finds the global minimum of the error sum of squares (the method yields maximum likelihood estimates when the disturbances are normal, independent, identically distributed variates). It proceeds as follows. First, 1b is estimated for the two smallest values of x greater than zero, and 1c is estimated for all other values of x ; when observations are available for $x = 0$, 1a is also estimated. Second, the point at which the two regression equations meet is computed. If this point lies in the closed interval from x_L to x_U , the solution is a join model; otherwise, it is a jump model. When the solution is a jump model, a join solution is obtained by estimating the parameters in 1b and 1c subject to the restriction that the regression lines meet at x_L . These steps are repeated after higher values of x are successively added to 1b and deleted from 1c. Thus, the program computes all possible solutions, out of which it prints the solutions of the join and jump model with minimum error sum of squares.

The program also computes and prints estimated asymptotic covariances and correlations of the parameter estimators. The estimates of the covariances are based on certain asymptotic properties of the parameter estimators (the asymptotics refer to an increase of the number of values of x without limit; if the values of x are not randomly sampled, the increase should be uniform between the experimentally set values of x). These properties are as follows: (1) The parameter estimators are consistent

and asymptotically normally distributed; (2) asymptotically, the parameter estimators of 1b are independent of the parameter estimators of 1c; (3) the asymptotic distribution of the parameter estimators in the join model does not differ from that in a model without continuity restriction for which the change point is known. These properties are proven by Feder (1975a) for models that are continuous in the change point, and by Van de Geer (1988, chap. 5) for a wider class of models, that includes models that are not continuous in the change point. Note, in particular, that the parameter estimators are not unbiased, for which Hinkley (1969) gives a simple reason.

The asymptotic properties imply that the asymptotic variances and covariances of the parameter estimators a_i and b_i are analogous to the estimators known from the small sample case. Thus, if $\mathbf{X}$ is the matrix that contains, in the appropriate order, the variables x and dummy variables for the constants in 1a–1c, then the asymptotic covariance matrix is estimated by $s^2(\mathbf{X}'\mathbf{X})^{-1}$, where s^2 is a consistent estimator of the variance of the disturbances. The program estimates s^2 by SSE/n , where SSE is the error sum of squares, and the divisor n is the number of observations instead of the degrees of freedom for error, which is the familiar small-sample divisor (the estimate of s^2 is identical to the maximum likelihood estimate in the small-sample case if the disturbances are assumed to be normal, independent, identically distributed variates).

For the model with a discrete independent variable, the estimates for the join model are known to satisfy the restriction that the meet should occur at one of a limited set of x values. Such restrictions should therefore also be taken into account in estimating the asymptotic covariance matrix of the parameter estimators. Suppose that the estimated meet for the model occurs at $x = c$. The estimates then satisfy the restriction $a_1 + b_1 c = a_2 + b_2 c$, or, equivalently, $a_1 + b_1 c - a_2 - b_2 c = 0$. If $\mathbf{b}$ is a vector which contains the parameters and $\mathbf{k}'$ is the vector $(1 \ c - 1 \ -c)$, then the restriction can be written as $\mathbf{k}'\mathbf{b} = 0$. The asymptotic covariance matrix of the estimators is assumed to have the same form as that of the small-sample case, and is therefore estimated by

$$s^2(\mathbf{X}'\mathbf{X})^{-1} - s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{k}[\mathbf{k}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{k}]^{-1}\mathbf{k}'(\mathbf{X}'\mathbf{X})^{-1}$$

(see, e.g., Judge, Griffiths, Hill, Lütkepohl, & Lee, 1985, p. 53). This estimator is not based on a rigorous proof, but the conditions that imply the properties of the estimators for the case of a continuous independent variable suggest that this estimator is consistent. In particular, the parameter estimators for the model with a discrete independent variable will still be consistent, asymptotically normally distributed, and the same, asymptotically, as for a model with a known change point. Of course, the properties of the estimators may not be completely identical to the properties of the estimators in the model with a continuous independent variable. For example, the rate of convergence of the estimators to their asymptotic distribution may be different from the convergence rate of the estimators in the model with a continuous independent

variable (e.g., the bias in small samples may be larger in the discrete case).

The program also prints the asymptotic standard error of the meet of the join model. If $\mathbf{d}'$ is the row vector that contains the partial derivatives of the estimate of the meet, $(a_2 - a_1) / (b_1 - b_2)$, with respect to the estimated parameters a_1, b_1, a_2, b_2 , respectively, and $\mathbf{C}$ is the estimated covariance matrix of the estimated parameters, then the asymptotic variance of the meet is estimated as $\mathbf{d}'\mathbf{C}\mathbf{d}$ (Feder, 1975a; Van de Geer, 1988). Of course, this estimator is not accurate for a model with a discrete independent variable. There is some evidence that the asymptotic normal distribution with variance $\mathbf{d}'\mathbf{C}\mathbf{d}$ may not be a very good approximation for small sample sizes (cf. Hinkley, 1969). Unfortunately, such a simple estimator for the asymptotic variance does not seem to exist for the change point of a jump model. Therefore, the program does not give the standard error for the change point of the jump model.

SOME PROGRAM CHARACTERISTICS AND OPTIONS

The program is written in Turbo Pascal and runs on an MS-DOS IBM-compatible personal computer. It should be fairly easy to transport the program to other machines. Input data consist of pairs of x, y values in free format. Data from a group of subjects can be analyzed in one run. There is no restriction on the number of subjects. Since memory for each subject is allocated dynamically, the amount of data per subject is limited only by the amount of memory available on the computer. The program reads its instructions (e.g., about options and input and output files) from a file and prints solutions for both the join and the jump model. A program is available that can create the input file interactively. The output consists of the parameter estimates, the meet, their estimated asymptotic standard errors, the lower bound and the upper bound of the interval in which the change occurs from 1b to 1c, the error sum of squares SSE, nine measures of fit for both jump and join models, and the estimated asymptotic variances, covariances, and correlations of the parameter estimates. Among the measures of fit are R^2 , the squared correlation between predicted and observed values of y , and a goodness-of-fit measure called SSTMFIT, and defined as $1 - \text{SSTM} / \text{SSE}$, where SSTM is the sum of squares around the mean of y . These measures of fit are often equal, but they need not be when linear restrictions are imposed on the parameters. In addition, measures of fit adjusted for the number of estimated parameters are reported.

SSTMFIT measures the reduction in error sum of squares after fitting a general constant. This is an appropriate measure for the type of regression models that are commonly fitted, because often no explanatory value can be attributed to the fact that the general constant is not equal to zero. However, 1b and 1c contain two constants, and therefore, a measure of fit is also included

that assesses fit due to regression weights only (i.e., after fitting the constants). Whether this measure is useful depends on the explanatory value of the constants. The measures of fit and their motivation are described in more detail in the program manual (Beem, 1993).

The program Segcurve has a number of options, which can be invoked by setting numbers to the required values. The manual contains detailed information about the options, but some of them will be described here briefly.

Weighted Least Squares

The program has an option to compute a weighted least squares solution based upon weights provided by the user. This is useful in at least two cases. First, if there are many observations at the levels of x , computation time can be reduced by analyzing means instead of raw scores. If the number of observations is not the same for each level of x , the variance of the means will, in general, be different, and therefore a weighted least squares solution may be preferred. Second, the variance of the raw scores may be different for different levels of x . For example, when response latencies are used as measures of processing effort, the mean and variance are often positively related. As another example, proportion correct, p , may be used as the dependent variable, and in that case, $p(1 - p)$ can be used as an estimate of the variance (alternatively, of course, a transformation may be used, such as the arcsine transformation or the logarithm of the odds ratios).

Investigating Hypotheses

Parameters can be estimated, subject to any linear restriction that the user may wish to make. The user should supply the relevant contrast vectors for the restrictions. For example, for mental rotation, the following hypotheses may be of interest:

$$a_0 = a_1 \quad (1)$$

$$a_0 = a_2 + b_2 * 360 \quad (2)$$

$$a_1 = a_2 + b_2 * 360 \quad (3)$$

$$a_0 = a_1 = a_2 + b_2 * 360 \quad (4)$$

The value 360 is of course specific to mental-rotation data. The third hypothesis corresponds to equal latencies for all processes, other than rotation, under the two strategies. The estimated covariance matrix for restricted parameters has the same form as the covariance matrix for the discrete model, except that if more than one linear contrast is provided, the vector $\mathbf{k}'$ is replaced by a matrix, $\mathbf{K}'$.

To take an example from another domain, consider Siegler's (1987) proposal that the strategies suggested by Woods et al. (1977) do not really occur. He suggested that children always count down, but keep track of whether they have exhausted the smaller number and of whether they have reached the smaller number. If the present model were fitted, Siegler's interpretation im-

plies the null hypothesis that the change point always occurs at the same (optimal) point for all subjects and that the sum of the regression weights under the two strategies is zero for all subjects. Thus, Siegler's proposal can be formulated as a linear hypothesis about the parameters of the two-phase model.

Finally, the linear restrictions can serve the purpose of fitting special cases of the model 1a–1c. The models fitted by Mathews et al. (1988) imply that for the initial learning phase, b_1 should be set to zero, whereas for the phase of asymptotic performance, b_2 should be set to zero and a_2 to the maximal performance score.

Since the program's output contains the measures of fit, hypotheses can be investigated by comparing the measures of fit under various restrictions on the parameters. Alternatively, the companion program SegcStat can be used to test hypotheses formally by applying significance tests. Two test statistics and their descriptive significance levels are provided. If SSE (H1) and SSE (H0) are the error sum of squares under hypotheses H1 and H0, respectively, where H0 is nested in H1 (i.e., H1 is the least, and H0 the most, restrictive hypothesis), and n is the number of observations, the first statistic then has the form $n * [\text{SSE}(\text{H0}) - \text{SSE}(\text{H1})] / \text{SSE}(\text{H1})$. This statistic has an asymptotic chi-squared distribution on q degrees of freedom, where q is the difference between the number of restrictions under H0 and the number under H1 (Van de Geer, 1988). The second statistic has the form $n * 1n [\text{SSE}(\text{H0}) / \text{SSE}(\text{H1})]$. This statistic has the same asymptotic distribution as the first (Feder, 1975b). The corresponding F statistics and their descriptive significance levels are also printed. These may be better approximate test statistics in finite samples, but there is no guarantee that this is actually the case.

Specific tests for a jump model versus a join model, or for a one-phase model versus a two-phase model, are not included in the program. The statistics appropriate for such tests are difficult to compute and converge rather slowly to their asymptotic distribution (e.g., Feder, 1975b). Segcurve optionally computes and writes to a file statistics that can be used to test for a one-phase model versus a wider class of alternative models. The statistics are the forward and backward cumulative sums of the so-called recursive residuals. In order to use the cumulative sums as test statistics, boundaries must be determined that define the critical regions of the test. Because no uniformly optimal boundaries exist for this kind of test, a choice must be made among many—in fact, infinitely many—boundaries. The program computes the approximate critical regions for the cumulative sums proposed by Brown, Durbin, and Evans (1975), but users can use the output to compute other test statistics and critical regions (see Buckley, 1991; Deshayes & Picard, 1982; Schweder, 1976). The program manual contains further information.

SegcStat can also be used to produce condensed output. This is useful if the output of Segcurve is considered

too lengthy when the data for a large number of subjects are analyzed in one run. Segcurve can be instructed to create a file from which SegcStat can produce condensed output.

AN EXAMPLE FROM A MENTAL-ROTATION EXPERIMENT

The parameters of 1b and 1c are often estimated by means of a nonlinear least squares estimation method. In that case, the join model is reformulated to ensure that the model is continuous in the change point. For example, 1b is written as before, and 1c is written as $y_{ij} = a_1 + b_1s + b_2(x_{ij} - s) + e$. This formulation demonstrates that the model is nonlinear in the parameters. As is often the case with nonlinear minimization methods, these methods are not guaranteed to find the best-fitting model in a least squares sense. A further problem arises with this model because the derivative at $x_i = s$ does not exist. As a consequence, minimization methods based on first derivatives are theoretically not appropriate. In practice, methods based on numerical derivatives will often do fairly well. However, if the stopping criterion is too stringent, the method may not converge (if the change point is approximated very closely at some iteration, the method may jump over it in the next iteration). I am not aware of a method to estimate the jump model by means of a nonlinear minimization method.

The model was applied to data from a mental-rotation experiment conducted by Rust (1988). Response latencies were analyzed with the program Segcurve and with the procedure CNLR in SPSSX 3.0. The estimates of the meet (i.e., the change points) of the two programs for the join model with a continuous independent variable will be compared.

In the experiment, a trial consisted of the presentation of two figures, which differed in their orientation angle. The difference in orientation ranged from 20° to 340° of angular separation, increasing in steps of 20°. Subjects were asked to state quickly and accurately whether the figures were the same or different. In positive pairs, the figures were identical, whereas in negative pairs, they were mirror images. Ten positive and ten negative pairs were presented at each orientation angle. Of the 63 subjects who took part, the data of five were removed from the analysis because they were not well fitted by the model.

Table 1
Summary Statistics for Absolute Value of Difference Between Segcurve and CNLR Meet Estimates

Measure	Unrestricted Estimates		Restricted Estimates	
	Raw Data	Median	Raw Data	Median
Mean	46	50	90	114
Minimum	4	7	2	7
Maximum	135	111	407	442
SD	34	36	121	135
N	19	15	15	16

Note—Statistics are for subjects with different meet estimates from CNLR and Segcurve. SD, standard deviation; N, number of subjects.

Table 2
Segcurve and CNLR Meet Estimates, SSTMFIT, and Segcurve Standard Error of Meet for Restricted Estimates for Medians

Subject Number	CNLR		Segcurve		SE
	Meet	SSTMFIT	Meet	SSTMFIT	
22	126	.76	114	.76	10.70
24	510	.12	160	.57	17.30
27	622	.22	180	.66	14.80
29	164	.62	129	.66	13.75
31	138	.70	145	.71	12.49
32	148	.76	125	.78	9.84
33	157	.87	166	.87	7.59
34	97	.56	174	.65	14.79
35	4	.00	167	.81	9.51
38	142	.43	180	.44	23.22
43	358	.00	133	.62	15.74
44	18	.02	171	.76	11.33
56	383	.00	158	.55	18.28
57	150	.79	125	.81	9.08
62	148	.62	176	.63	15.61
65	122	.88	114	.88	7.08

Note—SSTMFIT = $1 - \text{SSE} / \text{SSTM}$, where SSE and SSTM are, respectively, the error sum of squares and the total sum of squares around the mean; SE, estimated standard error of meet from Segcurve.

Raw latency scores and median latencies at each level of x were analyzed for positive pairs. The starting values for the CNLR procedure were the same for all subjects; for the change point, the value was 180°. Restricted estimates were also computed. These were constrained to satisfy the linear hypothesis $a_1 = a_2 + b_2 * 360$ (this hypothesis was rejected at the .05 level for a few subjects only). The jump model appeared to fit substantially better for only three subjects.

Table 1 contains summary statistics for the absolute value of the difference between the Segcurve and CNLR meet estimates for those subjects for whom Segcurve and CNLR produced different meet estimates for the join model. The statistics demonstrate that the differences between the estimates—especially the restricted estimates—of the meet can be quite substantial. The meet estimates are different for approximately 25–35 percent of the subjects. Table 2 contains the meet estimates of the restricted model and the values of SSTMFIT for those subjects for whom the Segcurve and CNLR meet estimates for median latencies differed. The estimated standard error of the meet estimate from Segcurve is also presented. The differences between the Segcurve and CNLR meet estimates are within two times the estimated standard error for only five subjects. Evidently, the estimates of the nonlinear estimation procedure can be widely different from the optimal estimate provided by Segcurve. This is true even when SSTMFIT, the measure of fit provided by CNLR, is not very different from SSTMFIT for the optimal solution.

Program Availability

The compiled program, the source code, and a manual are available from the author, from whom information can be obtained by writing or via e-mail.

REFERENCES

- BEEM, A. L. (1993). *Segcurve: A program for fitting two-phase segmented curve models with an unknown change point* [Program manual]. Leiden, The Netherlands: Leiden University, Centre for the Study of Education and Instruction.
- BROWN, R. L., DURBIN, J., & EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B*, **37**, 149-163.
- BUCKLEY, M. J. (1991). Detecting a smooth signal: Optimality of cusum based procedures. *Biometrika*, **78**, 253-262.
- COOPER, L. A. (1980). Spatial information processing: Strategies for research. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction* (Vol. 1, pp. 149-176). Hillsdale, NJ: Erlbaum.
- DESHAYES, J., & PICARD, D. (1982). Tests of disorder of regression: Asymptotic comparison. *Theory of Probability & Its Applications*, **27**, 100-115.
- FEDER, P. I. (1975a). On asymptotic distribution theory in segmented regression problems—identified case. *Annals of Statistics*, **3**, 49-83.
- FEDER, P. I. (1975b). The log likelihood ratio in segmented regression. *Annals of Statistics*, **3**, 84-97.
- GLUSHKO, R. J., & COOPER, L. A. (1978). Spatial comprehension and comparison processes in verification tasks. *Cognitive Psychology*, **10**, 391-421.
- HINKLEY, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, **56**, 635-639.
- HINRICH, J. V., YURKO, D. S., & HU, J.-M. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 890-901.
- HUDSON, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association*, **61**, 1097-1129.
- IPPEL, M. J., & BEEM, A. L. (1987). A theory of antagonistic strategies. In E. De Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction: European research in an international context* (pp. 111-121). Oxford: Pergamon.
- JUDGE, G. C., GRIFFITHS, W. E., HILL, R. C., LÜTKEPOHL, H., & LEE, T.-C. (1985). *The theory and practice of econometrics*. New York: Wiley.
- KYLLONEN, P. C., LOHMAN, D. F., & SNOW, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology*, **76**, 130-145.
- KYLLONEN, P. C., LOHMAN, D. F., & WOLTZ, D. J. (1984). Componential modelling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology*, **76**, 1325-1345.
- LOHMAN, D. F., & IPPEL, M. J. (1993). Cognitive diagnosis: From statistically based assessment toward theory-based assessment. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 41-71). Hillsdale, NJ: Erlbaum.
- MATHEWS, R. C., BUSS, R. R., CHINN, R., & STANLEY, W. B. (1988). The role of implicit learning processes in concept discovery. *Quarterly Journal of Experimental Psychology*, **40A**, 135-165.
- RUST, T. R. (1988). *Strategy shifting on a spatial rotation task*. Unpublished master's thesis, University of Iowa, Iowa City.
- SCHWEDER, T. (1976). Some "optimal" methods to detect structural shift or outliers in regression. *Journal of the American Statistical Association*, **71**, 491-501.
- SIEGLER, R. S. (1987). Strategy choices in subtraction. In J. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 81-106). Oxford: Oxford University Press.
- SNOW, R. E. (1980). Aptitude processes. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction* (Vol. 1, pp. 27-64). Hillsdale, NJ: Erlbaum.
- SNOW, R. E., & LOHMAN, D. F. (1993). Cognitive psychology, new test design, and new test theory: An introduction. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 1-17). Hillsdale, NJ: Erlbaum.
- VAN DE GEER, S. A. (1988). *Regression analysis and empirical processes* (Centrum voor Wiskunde en Informatica Tract 45). Amsterdam: Centre of Mathematics and Computer Science.
- WOODS, S. S., RESNICK, L. B., & GROEN, G. (1977). An experimental test of five process models for subtraction. *Journal of Experimental Psychology*, **102**, 734-736.

(Manuscript received May 21, 1993;
revision accepted for publication June 24, 1994.)