

Using segmented linear regression models with unknown change points to analyze strategy shifts in cognitive tasks

KOEN LUWEL

University of Leuven, Leuven, Belgium

A. LEO BEEM

Free University of Amsterdam, Amsterdam, The Netherlands

and

PATRICK ONGHENA and LIEVEN VERSCHAFFEL

University of Leuven, Leuven, Belgium

Some years ago, Beem (1993, 1995) described a program for fitting two regression lines with an unknown change point (Segcurve). He suggested that such models are useful for the analysis of a variety of phenomena and gave an example of an application to the study of strategy shifts in a mental rotation task. This technique has also proven to be very fruitful for investigating strategy use and strategy shifts in other cognitive tasks. Recently, Beem (1999) developed SegcurvN, which fits n regression lines with $(n-1)$ unknown change points. In the present article we present this new technique and demonstrate the usefulness of a three-phase segmented linear regression model for the identification of strategies and strategy shifts in cognitive tasks by applying it to data from a numerosity judgment experiment. The advantages and shortcomings of this technique are evaluated.

One of the major methodological problems in developmental psychology and the study of individual differences is to determine the strategy used by a subject for a particular range of items in a given task. The determination of strategy use is important, because in most current theories of strategic competence, like the one of Lemaire and Siegler (1995), task performance is a function not only of the efficiency with which the available strategies are executed, but also of (developmental) changes in strategy use and of the frequency and the adaptiveness with which the available strategies are executed. The tasks may come from a variety of domains, such as simple addition (Siegler & Robinson, 1982) or subtraction (Siegler, 1987), serial recall (McGilly & Siegler, 1990), spelling (Marsh, Friedman, Welch, & Desberg, 1980), mental rotation (Ippel & Beem, 1987), and others.

Thus there is a great need for techniques of data gathering and for statistical models that allow a fine-grained analysis of what strategy a subject has applied on different items of the same task. In some studies retrospective reports have been used as the primary source of data about strategy use. However, in many cases it is not possible or

warranted to rely (exclusively) on that kind of data, and alternative (or additional) data-gathering techniques, like eye-fixation records, accuracy measures, and response times, are required to get (complementary) information about the strategies underlying subjects' task performance.

A typical characteristic of these latter kinds of data is that, although they yield some information about the manner in which a subject has approached a particular item, this information does not lead unambiguously and directly to the identification of the underlying strategy. Nevertheless, the pattern in the data (e.g., number of eye fixations, deviation from the correct answer, response time) over the whole range of items may provide valuable information about what strategies a subject has used for the task as a whole, how the use of different strategies is related to certain item characteristics, and, last but not least, where strategy shifts occur.

In this article we will describe and evaluate the application of segmented linear regression models to investigate phenomena of strategy use and strategy shifts in data patterns in which there is a linear relationship between a task parameter (e.g., numerosity or angle of rotation) and a dependent variable (e.g., response time or a measure of accuracy). The models are implemented in a computer program called SegcurvN (Beem, 1999), which fits segmented linear regression models with an arbitrary number of segments to individuals' response-time or error-rate patterns. The linearity is a limitation of the program not of segmented curve models in general. SegcurvN is an elabora-

This study was supported by Grant G. 0157.98 from the Fund of Scientific Research. The authors thank Richard Carlson for his valuable comments on a previous version of this article. Correspondence should be addressed to K. Luwel, Center for Instructional Psychology and Technology, University of Leuven, Vesaliusstraat 2, B-3000 Leuven, Belgium (e-mail: koen.luwel@ped.kuleuven.ac.be).

tion of an earlier program, Segcurve (Beem, 1993, 1995), which fits two-phase segmented linear regression models to a subject's data pattern. To set the stage, the next section discusses some alternative methods that have been previously used to analyze and detect strategy shifts.

ALTERNATIVE METHODS FOR ANALYZING AND DETECTING STRATEGY SHIFTS

As noted, the identification of strategies and strategy shifts in a cognitive task has proven to be a difficult issue. We will briefly describe two ways in which strategies were identified in previous research. The first procedure consists of dividing the range of the independent variable into different segments and comparing the means of the dependent measures in each segment for distinct groups of subjects. In the second—more advanced—technique, a regression model of which the different segments identify the different strategies is fitted on subjects' response-time or error-rate patterns.

The first technique was applied by Newman and Berger (1984) in a study of children's strategy development for estimating ordinal numbers. Children of different age groups (kindergarten, first, and third grade) played a game of darts on a computer. The screen showed a vertical number line of which only the beginning (1) and the end point (23) were shown. At each trial, a randomly presented balloon appeared somewhere along the number line and the child had to guess the ordinal number represented by the balloon's position. After the guess was entered into the computer, a dart was sent horizontally across the screen, providing feedback about the accuracy of the child's guess. The children had to keep on sending darts until the guess was correct. After completion of the game, children were asked to explain briefly how they had handled the task.

On the basis of the assumption that children of different ages would differ in terms of the kinds of strategies applied for different kinds of items, Newman and Berger (1984) compared the estimation accuracy of the three age groups for small (2–8), medium (9–15), and large (16–23) target positions. Accuracy was measured in two ways: (1) the number of misses before hitting the target and (2) the mean absolute deviation of the initial estimate from the target. This comparison showed for both measures that both younger groups were more accurate in the small range relative to the medium and large range, whereas the third graders were more accurate at either of the two extremes relative to the medium range. Second, the verbal reports of children's strategy use revealed a clear development toward a greater flexibility in strategy use. The majority of kindergartners always reported counting up from 1 in each of the different ranges. First graders indicated counting up from 1 in the small range and counting down from 23 in the large range, and used one of both strategies in the medium range. The oldest children reported applying the same strategies in the small and large range as the first graders, but they started to count up from the middle (i.e.,

10 or 15) in the medium range. Children's explanations of strategy use were related to estimation accuracy.

For the data analysis, the number of segments in which the range of the independent variable is divided was chosen in an arbitrary way. As a consequence, the results of this technique can vary according to this a priori decision made by the researchers. Moreover, Newman and Berger's (1984) analysis did not allow a precise identification of the values of the independent variable (i.e., the size of the target) at which subjects shifted from the counting-up to the counting-down strategy. Instead, they assumed that the location of this change point was the same for all subjects.

A second and more advanced technique for identifying a subject's strategy use is fitting a regression model to the subject's response times or error rates. Kyllonen, Lohman, and Woltz (1984) fitted regression models to a subject's response times and error rates to study strategy shifts on different items of the same spatial task as a function of item characteristics. For each item in the task, three steps needed to be completed. First, during the encoding step, subjects were presented a geometrical figure which was referred to as the A figure. Next, during the synthesis step, the A figure was removed, and subjects were presented one or two other figures, the B and C figures, which they had to combine mentally with the A figure. Finally, during the comparison step, the B and C figures were removed and a test probe was presented. Subjects had to indicate whether the image that was mentally formed during the synthesis step was the same as the test probe. Three item characteristics were manipulated: (1) the number of figures to be combined with the A figure, (2) synthesis from the left or the right side of the B and C figures, and (3) test probe the same as the constructed image for half of the items and different for the other half. Each fitted regression 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 change 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^2 statistics from the 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 different subjects differ in the point at which they switch from one strategy to another.

Although Newman and Berger (1984) and Kyllonen et al. (1984) were able to identify different strategies that were used by the different subjects, the change points in both studies had to be determined a priori by the researcher. Moreover, the location of these change points was the same for all subjects.

In order to estimate the position of the change point for each individual separately, a regression model must incorporate the location of change as a parameter and estimate the parameter from the observed data of each subject. Such regression models are variously called segmented

curve models (Hudson, 1966) or switching regressions with unknown change or break points (Judge, Griffiths, Hill, Lütkepohl, & Lee, 1985). More generally, detecting and estimating the location or time of (abrupt) structural changes in (seemingly) nondeterministic systems is also referred to as change point analysis (Chen & Gupta, 2000; Csörgö & Horváth, 1997; Quandt, 1958). In the next section we will give a brief description of a two-phase segmented linear regression model followed by an application of this model on the data of a numerosity judgment task.

THE TWO-PHASE SEGMENTED LINEAR REGRESSION MODEL

The most simple segmented linear regression model contains one change point. Beem (1993, 1995) presented a program, *Segcurve*, by which such models can be fitted for each subject. More specifically, the program estimates two regression equations that hold for different ranges of the independent variable and estimates the change point for each subject separately. The data of all subjects can be fitted in one run of the program. In other words, this program is ideally suited for the study of subjects' strategy shifts in cognitive tasks with one unknown change point. The model fitted by the program can be formally described as follows:

$$y = a_1 + b_1x + e, \text{ for } x \leq s \quad (1)$$

$$y = a_2 + b_2x + e, \text{ for } x > s \quad (2)$$

where x is the independent and y is the dependent variable, the parameters a_i and b_i ($i = 1, 2$) denote respectively the intercept and the slope of the regression lines, and e is the error term. The parameter s is called the change point or break point. For values of the independent variable up to s the first regression equation is fitted, whereas for values larger than s the second equation—with a different intercept and slope—holds. The program simultaneously estimates the regression parameters and change point using a least squares algorithm that always finds the optimal solution.

Two variants of the model can be distinguished: the *join* model and the *jump* model. The join model is continuous in s . This means that the value of y at s is the same for both regression equations. When the model is not continuous in s , it is called a jump model. The variable x may be continuous, as in mental rotation, or it may be essentially discrete, as in numerosity judgments. Beem (1995) discussed some consequences of this distinction for the definition and estimation of a jump model and a join model. A more detailed description of the program's properties and the various statistics that it computes can be found in Beem (1993, 1995).

Ippel and Beem (1987) and Alderton and Larson (1994) applied this model to a mental rotation task. In a mental rotation task, subjects must decide 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. Either a clockwise or a counterclockwise

routine can achieve the rotation. Depending on the angular disparity of the figures and a subject's speed of clockwise and counterclockwise routine, one of these strategies is the most efficient. Ippel and Beem proposed that strategy shifts have as primary purpose reducing mental effort.

The model was also successfully applied in a study about the development of strategies for judging (cardinal) numerosities by Verschaffel, De Corte, Lamote, and Dhert (1998). In that study, subjects of three different age groups (i.e., university students, sixth graders, and second graders) had to determine the numerosity of 100 different amounts of colored square blocks presented in a 10×10 grid, whereby each trial was presented for at most 20 sec. As in the study of Newman and Berger (1984), it was assumed that this task allowed for the use of two different kinds of strategies, with each strategy leading to a predictably different course of response times for the range of items for which they were applied. First, in an addition strategy, the subject divides the given quantity of blocks into a number of groups, determines the number of blocks in each group, and adds this result to a running total. This strategy was assumed to result in linearly increasing response times with a growing number of blocks. Second, in a subtraction strategy, the numerosity of blocks is determined by subtracting the number of empty squares from the total number of squares in the grid (i.e., from 100). It was assumed that this strategy would elicit linearly decreasing response times with a growing number of blocks.

To identify subjects' strategy use, Verschaffel et al. (1998) fitted a one-phase and a two-phase regression model to the patterns of response times on all trials. The independent variable was the number of blocks in a trial, running from 1 to 100. The one-phase model corresponds to a hypothetical model wherein all items are solved exclusively by means of the addition strategy. The two-phase model corresponds to a hypothetical model wherein small and large numerosities are solved, respectively, by means of an addition and a subtraction strategy. Verschaffel et al. used the jump model for their analyses, since their rational task analysis does not necessarily suggest continuity at the change point.

Results provided support for the hypothesized developmental trend toward the greater (adaptive) use of the more efficient subtraction strategy: The fit of the two-phase segmented linear regression model increased with age, whereas the fit of the one-phase model decreased with age. Moreover, a strong and positive relationship was found between the (adaptive) use of the subtraction strategy (as evidenced by a high fit of subjects' response-time patterns to the two-phase segmented linear regression model), on the one hand, and their numerosity judgment accuracy (as indicated by the mean absolute deviation of the answer from the given numerosity in the 100 trials), on the other hand.

However, for a number of subjects in each age group the fit of the two-phase segmented linear regression model was rather low, due to the fact that the response times for the items in the middle region were not a continuation of those for the items with small or large numerosities. An

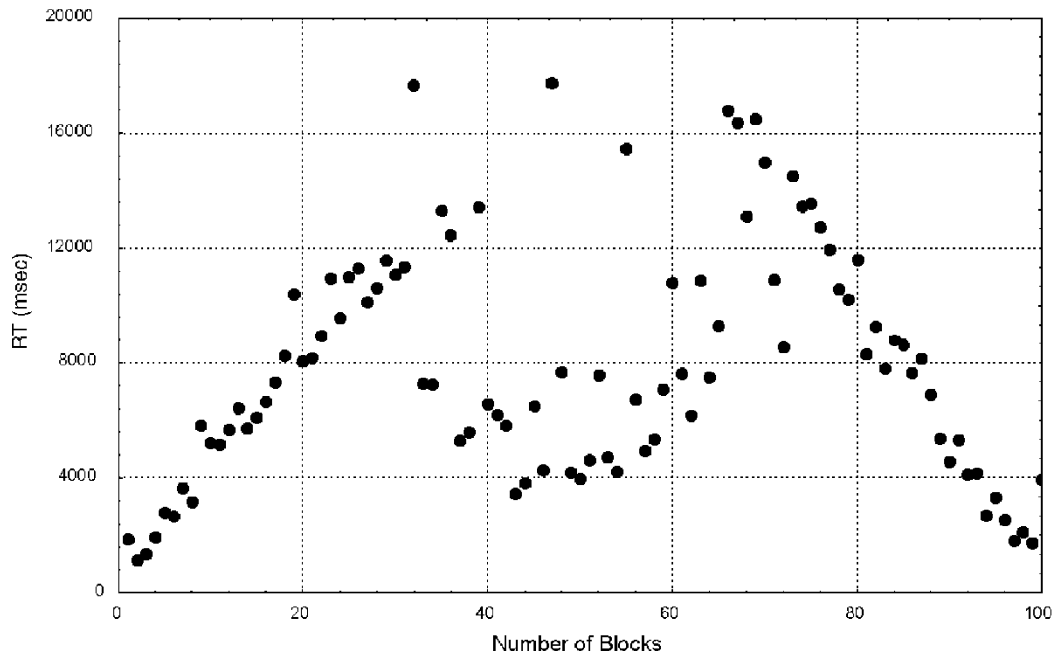


Figure 1. Example of a response-time pattern in which a third strategy was used for the trials in the middle range.

example of such an individual response-time pattern is given in Figure 1. A visual inspection of the response-time data of these subjects, in combination with their patterns of error rates, suggested that these subjects had used a third—unanticipated—strategy, the so-called estimation strategy. The basic characteristics of this estimation strategy are that it is relatively quick, that it leads to rather imprecise answers, and that its duration does not seem to be seriously affected by the numerosity to be determined.

The assumed occurrence of this estimation strategy for the items in the middle region of the independent variable jeopardized the use of the two-phase segmented linear regression model, which can identify only one shift between two strategies in a set of data points. As noted by Beem (1995, p. 393), segmented linear regression models can, in principle, be formulated for any number of strategies. In order to fit models with more than one change point, the program has been extended so that it can fit models with an arbitrary number of change points in any combination of joins and jumps for the change points. In the next section we will briefly describe this extension, thereby focusing on those aspects of the model that make it new in comparison with the previous one. For more (technical) information see the Appendix (see also Beem, 1999).

THE THREE-PHASE SEGMENTED LINEAR REGRESSION MODEL

The three-phase model is an extension of the two-phase model in which an additional change point and regression

equation, with additional parameters, is added to the two-phase model to account for a third solution strategy or type of behavior. The three-phase segmented linear regression model can be described as follows:

$$y = a_1 + b_1x + e, \text{ for } x \leq s_1 \tag{3}$$

$$y = a_2 + b_2x + e, \text{ for } s_1 < x \leq s_2 \tag{4}$$

$$y = a_3 + b_3x + e, \text{ for } x > s_2 \tag{5}$$

The relationship between the dependent and independent variable is described by three regression equations, which hold for different ranges of the independent variable. Equation 3 is fitted to all data points smaller than or equal to the first change point s_1 , Equation 4 is fitted to all data points larger than s_1 but smaller than or equal to s_2 , and Equation 5 is fitted to all data points larger than s_2 . The program fits the model under the assumption that the two change points s_1 and s_2 are unknown and thus must be estimated. Extension to an arbitrary number of change points and regression equations is straightforward. It merely requires adding more equations and change points to the model.

Whereas the two-phase segmented linear regression model has two variants, a *jump* and a *join* model, the three-phase model consists of four variants, a *join-join*, *join-jump*, *jump-join*, and *jump-jump* model, depending on whether the model is continuous or discontinuous in each of the change points.

We will illustrate the application of the three-phase segmented linear regression model by means of a study of

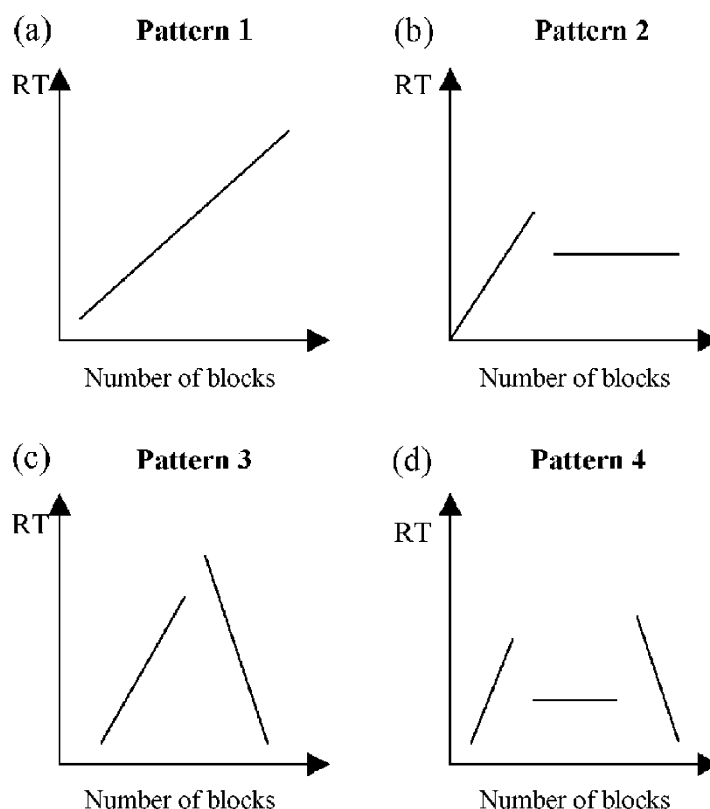


Figure 2. Hypothetical response-time patterns with (a) application of the addition strategy, (b) use of the addition and estimation strategy, (c) execution of the addition and subtraction strategy, and (d) application of the addition, estimation, and subtraction strategy.

Luwel, Verschaffel, Onghena, and De Corte (2001). This study deals with the same theoretical questions and uses the same methodology as the aforementioned study of Verschaffel et al. (1998). More specifically, subjects of two different age groups (i.e., second and sixth graders) were asked to judge numerosities of blocks that were presented in grids of three different sizes (i.e., 7×7 , 8×8 , and 9×9) for at most 20 sec. Luwel et al. (2001) hypothesized that subjects would use three instead of two strategies to solve this task—namely, an addition strategy, a subtraction strategy, and an estimation strategy. Since they had access to the version of the program that can fit segmented linear regression models with an arbitrary number of change points (Beem, 1999), they were able to take into account the response times produced by the estimation strategy. On the basis of an elaboration of the (partially empirically supported) rational task analysis of Verschaffel et al., on the one hand, and the assumed connection between the use of a particular strategy on a particular item and the time needed to solve that item, on the other hand, four hypothetical response-time patterns were generated (Figure 2). Graphs a and b describe, respectively, the pattern of the response times for subjects who use the addition strategy solely (Pattern 1) and for those who use it in combination with the estimation strategy (Pattern 2).

Graphs c and d show, respectively, the response-time pattern for subjects who use the addition and subtraction strategy (Pattern 3) and for persons who adaptively use the three numerosity judgment strategies involved in our rational task analysis (Pattern 4). For a detailed discussion of the hypotheses concerning the effect of different task and subject variables on the response-time patterns, we refer to Luwel et al. Hereafter we will describe in greater detail how we used the segmented linear regression models to identify the presence of the hypothetical data patterns in subjects' response-time patterns.

Defining the four hypothetical response-time patterns from Figure 2 in terms of the different parameters of the statistical models presented above leads to the following characterization of each pattern:

1. Pattern 1 (always addition): no change point, and the only b parameter is positive.
2. Pattern 2 (first addition, then estimation): one change point, a positive b_1 parameter, and a b_2 parameter with a value close to zero.
3. Pattern 3 (first addition, then subtraction): one change point, a positive b_1 parameter and a negative b_2 parameter.
4. Pattern 4 (first addition, then estimation, and finally subtraction): two change points, a positive b_1 parameter, a

b_2 parameter with a value close to zero, and a negative b_3 parameter.

Since Luwel et al.'s (2001) rational task analysis did not necessarily suggest continuity at the change point(s), they fitted the two-phase jump model and the three-phase jump-jump model to the individual response-time data.

In all analyses the independent variable (i.e., the number of blocks actually presented in the grid) was considered discrete. We started by fitting the three-phase jump-jump model on the response-time data of all subjects that were selected for the analysis. Next, the data were tested for the presence of two change points following the *cusums* method (Brown, Durbin, & Evans, 1975; Schweder, 1976), explained in the Appendix. When the cusum test yielded a change point in both parts of the split data pattern, the presence of a three-phase model was confirmed in that particular response-time pattern.

When the cusum test did not reveal two change points in the response-time data pattern of a given subject, a two-phase segmented jump model was fitted to these data. Next, a cusum test for a single change point was executed. If this test revealed the presence of a single change point, the particular data pattern consisted of two phases. When the cusum test for the two-phase jump model did not yield any evidence for a change point, we decided that the data fitted Pattern 1.

Subjects for whom the cusum test yielded two change points in their data pattern were further tested for Pattern 4. For this pattern, three distinct criteria needed to be met: $b_1 =$ significantly positive, $b_2 =$ close to zero, and $b_3 =$ significantly negative. For each criterion, the null hypothesis $H_0: b = 0$ was tested by running the analysis again, but this time for a restricted three-phase jump-jump model in which the value of the b parameter under consideration was set to 0. Next, for each of the three criteria, H_0 was compared with H_1 using the F -type statistic (see Appendix or Beem, 1993, 1999). If the null hypothesis was rejected, the value of the estimated b parameter in the unrestricted model was accepted. If this null hypothesis was not rejected, the b parameter was assumed not to differ significantly from 0.¹

The data patterns for which the cusum test yielded one change point were tested for Patterns 2 and 3. For both patterns, we tested whether the b_1 parameter was positive by following the same logic as explained in the previous paragraph. Next, we distinguished between both patterns by comparing a restricted two-phase jump model in which the value of the b_2 parameter was set to zero with the unrestricted two-phase jump model. When the null hypothesis of the restricted model was not rejected, we assumed that the estimation strategy was applied for that segment and thus that the data fitted Pattern 2. When the null hypothesis of $b_2 = 0$ was rejected, we looked at the value of the estimated b_2 parameter in the unrestricted model. If this value was negative, we concluded that the data fitted Pattern 3.

All response-time data patterns that did not fit the hypothetical data patterns were inspected for ceiling effects

or for outliers that could have a strong but artificial influence on the estimated b parameters. With respect to the ceiling effects, all trials with a response time of 20 sec were deleted to rule out their disturbing influence. With respect to the outliers, we used the Cook's D statistic (Myers, 1990; Neter, Kutner, Nachtsheim, & Wasserman, 1996) in order to determine the extent of their influence. More specifically, if the Cook's D value of the outlier was the only value that differed relatively strongly from the Cook's D values of the other observations within the same segment, it was considered as an influential outlier and therefore it was removed from the data set. Next, the aforementioned analysis was carried out again on the adapted data sets.

Averaged over the different grid sizes, 68% of the second graders and 85% of the sixth graders fitted one of the four hypothetical patterns. The analysis with the segmented linear regression models of these response-time patterns revealed that all of them fitted with Pattern 3 or 4. So, not only the response-time patterns of the sixth-graders but also those of the second graders showed clear evidence of the (adaptive) use of the sophisticated subtraction strategy.

Second, the results showed a clear interaction between age and grid size. Whereas the majority of the second graders fitted Pattern 4 for all three grid sizes, the distribution of the patterns of the sixth graders was heavily affected by grid size in the sense that the percentage of fits with Pattern 4 increased and the percentage of fits with Pattern 3 decreased with growing grid size. This indicates that whereas the second graders felt back on the faster but less accurate estimation strategy for all three grid sizes, the sixth graders resorted to this strategy only for the largest grid size.

Third, the segmented linear regression models allowed us to also test several hypotheses concerning the location of the change points in the response-time patterns. For instance, we found that for both Patterns 3 and 4 and for both age groups, the first segment was larger than the last one, indicating that, in line with our rational task analysis, the (simpler) addition strategy was applied on a larger number of trials than the (more complex) subtraction strategy. Moreover, we found that for Type 4 patterns, the length of

Table 1
Proportion of Second and Sixth Graders With Response-Time Patterns Corresponding to the Hypothetical Patterns of Each of the Grid Sizes

Grid Size	Pattern							
	1		2		3		4	
	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>
Second Graders								
7 × 7	0.00	0	0.00	0	0.23	3	0.77	10
8 × 8	0.00	0	0.06	1	0.13	2	0.81	13
9 × 9	0.08	1	0.00	0	0.08	1	0.84	10
Sixth Graders								
7 × 7	0.00	0	0.00	0	0.71	10	0.29	4
8 × 8	0.00	0	0.00	0	0.50	9	0.50	9
9 × 9	0.00	0	0.00	0	0.26	5	0.74	14

the first and the last segments remained stable whereas (only) the length of the middle segment increased with increasing grid size. This suggested that the estimation strategy is indeed a coping strategy, which is resorted to when the task demands no longer allow the use of more accurate but also more demanding strategies.

Finally, the segmented linear regression models also allowed an investigation of the relation between the use of the three kinds of strategies and their accuracies by computing the mean error rates for each segment of the response-time pattern. The analysis of the response-time patterns that fitted Pattern 3 or 4 showed that there was no significant difference in mean error rates produced by the addition strategy, on the one hand, and the subtraction strategy, on the other, although the former strategy was applied on the smallest numerosities and the latter strategy was applied on the largest numerosities from the item set. This was in line with the rational task analysis. Moreover, for the subjects with a response-time pattern fitting Pattern 4, the mean error rates for the judgments from the middle segment were, as expected, much larger than the error rates for the judgments for the first and the last segments, indicating that the estimation strategy was indeed a rather imprecise strategy.

DISCUSSION

Some years ago, Beem (1995) proposed segmented linear regression models as a versatile general class of models for investigating strategy use and strategy shifts in cognitive tasks. Ippel and Beem (1987) and Verschaffel et al. (1998) successfully applied a two-phase segmented linear regression model to a mental rotation and a numerosity judgment task, respectively. The study described here demonstrates that a three-phase model can be useful for investigating the development of children's strategies for numerosity judgment in square grids.

We have argued that segmented linear regression models are more appropriate and more informative for studying strategy shifts in cognitive tasks than other kinds of analytic techniques. First, these models seem far more appropriate than the approach followed by Newman and Berger (1984) in their investigation of young children's strategies for determining ordinal numbers. As we already mentioned, they arbitrarily divided the number range into three segments. Their approach has the inherent risk that the outcome of the analyses depends on the number (and the boundaries) of segments in which the range of the independent variable is divided. Moreover, it is assumed that the location of the change points is the same for all subjects.

These problems are avoided when segmented linear regression models are used for which the change points are estimated for each subject separately instead of assuming a priori some specific change points at a certain location in the range. The method of Newman and Berger (1984) may artificially suggest abrupt developmental changes for age groups or for subjects measured at different points in time. In contrast, the present method would reveal how

much variation exists in each age group, which can obviously be relevant for the description of developmental processes.

Second, segmented linear regression models are also more appropriate and more informative for studying strategy shifts in cognitive tasks than quadratic or cubic models (Myers, 1990; Neter et al., 1996). The change points in these latter models are represented by the one or two "top" values that are computed by the models. However, the use of these models is restricted to situations in which the strategy shift is assumed to be continuous in the change point. The segmented linear regression models have a broader applicability relative to the quadratic and cubic models, since the join and jump variants allow the detection of strategy changes that are continuous as well as discontinuous in the change points. The model parameters are also more easily related to psychological processes such as the speed of counting. Notwithstanding this advantage, further research in which (the results of) both types of models are systematically compared might be interesting. An interesting question in this respect would be whether both types of models compute, respectively, the same number of top values or change points for a given data pattern, and whether the location of these top values and change points would be the same.

The reported study shows that the three-phase model allowed us to identify, characterize, and delineate—besides the addition and subtraction strategy already identified by Verschaffel et al. (1998)—a third, qualitatively different, strategy for numerosity judgment, the so-called estimation strategy. Moreover, the three-phase model also allowed a statistical analysis of subjects' use of and shift toward the subtraction strategy in response-time patterns fitting Pattern 4, which was impossible to realize with the two-phase segmented linear regression model.

Some problems were also apparent in the application of the model. First, when the number of observations within a particular segment was small, the statistical power of the statistics for testing linear hypotheses about the regression parameters was too small. For instance, when a specific segment with a large value of the b parameter consists of merely 10 data points, the tests may not reject the null hypothesis that $b = 0$. However, this is not so much a property of the test as due to the inherent noise in reaction times of children. The most obvious remedy is obtaining more observations.

Second, detecting strategy shifts by formal significance tests can be quite difficult if the shift occurs early in the beginning or near the end of the data pattern. Verschaffel et al. (1998) and Luwel et al. (2001) confirmed this well-known problem in change point analysis. Their visual inspection of subjects' response times and error rates strongly suggested use of the subtraction strategy on at least some of the items with the largest numerosities. However, the statistical analysis did not identify this subject as a user of the subtraction strategy (= Pattern 3 or 4 in Figure 2).

The cusum procedure has some advantages for the detection of change points because it is also a graphical method

and comparatively model free. However, a relatively recent growing interest in change point analysis has resulted in an improved theoretical understanding of other change point testing methods. Csörgő and Horváth (1997) have presented the asymptotic distribution theory of several test statistics in considerable detail. Among these is the log likelihood ratio (LR) test. The log LR test may be particularly relevant when strategies are qualitatively different because it can be used to formally test whether the variances of the disturbances e in Equations 3–5 are different for different strategies. Bai (1999) has also discussed the log likelihood ratio test and provided some Monte Carlo results. Chen and Gupta (2000) have advocated information theoretic procedures. Many of these statistics can be computed from the output that SegcurvN can write to a file for use by other programs.

As noted by Beem (1995), the analysis of strategy shifting by segmented linear regression models also provides an opportunity to assess, in a quite general sense, the optimality or “adaptiveness” (Lemaire & Siegler, 1995) of a subject’s performance, given the resources available to the subject. Newell (1990) has proposed that the degree of optimality (or adaptiveness) given available resources may be used as a *definition* of intelligent task performance. For mental rotation, optimal performance given a subject’s proficiency in clockwise and counterclockwise rotation implies a join model. A join model can provide a better fit than a jump model for a continuous independent variable. This is not true for a discrete independent variable. In that case, the degree of optimality can be measured as the difference between the estimated expected response times under the two strategies at the point where the subject’s strategy shift occurs. When strategies are qualitatively different, as in the numerosity judgment task, optimality may be investigated by evaluating some function of speed and accuracy for different shift points given the estimated parameters for the different strategies.

Program Availability

The compiled program, the source code, and a manual are available from A.L.B. by e-mail at al.beem@psy.vu.nl or by regular mail at Department of Biological Psychology, Free University, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.

REFERENCES

- ALDERTON, D. L., & LARSON, G. E. (1994). Cross-task consistency in strategy use and the relationship with intelligence. *Intelligence*, **18**, 47-76.
- BAI, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, **91**, 299-323.
- 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, Educational Computing Research Unit.
- BEEM, A. L. (1995). 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. *Behavior Research Methods, Instruments, & Computers*, **27**, 392-399.
- BEEM, A. L. (1999). *SegcurvN: A program for fitting segmented curve models with unknown change points* [Program manual]. Amsterdam: Free University, Department of Biological Psychology.
- 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.
- CHEN, J., & GUPTA, A. K. (2000). *Parametric change point analysis*. Boston: Birkhäuser.
- CÖRÖGÖ, M., & HORVÁTH, L. (1997). *Limit theorems in change-point analysis*. Chichester, U.K.: Wiley.
- 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* (Vol. 1, pp. 111-121). Leuven: Leuven University Press.
- JUDGE, G. C., GRIFFITHS, W. E., HILL, R. C., LÜTKEPOHL, H., & LEE, T.-C. (1985). *The theory and practice of econometrics*. New York: Wiley.
- KYLLÖNEN, P. C., LÖHMAN, D. F., & WOLTZ, D. J. (1984). Componential modelling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology*, **76**, 1325-1345.
- LEMAIRE, P., & SIEGLER, R. S. (1995). Four aspects of strategic change: Contributions to children’s learning of multiplication. *Journal of Experimental Psychology: General*, **124**, 83-97.
- L UWEL, K., VERSCHAFFEL, L., ONGHENA, P., & DE CORTE, E. (2001). Strategic aspects of children’s numerosity judgement. *European Journal of Psychology of Education*, **17**, 233-255.
- MARSH, G., FRIEDMAN, M., WELCH, V., & DESBERG, P. (1980). The development of strategies in spelling. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 339-354). London: Academic Press.
- MCGILLY, K., & SIEGLER, R. S. (1990). The influence of encoding and strategic knowledge on children’s choices among serial recall strategies. *Developmental Psychology*, **26**, 931-941.
- MYERS, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). Boston: PWS-Kent.
- NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., & WASSERMAN, W. (1996). *Applied linear statistical models* (4th ed.). Chicago: Irwin.
- NEWELL, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- NEWMAN, S. R., & BERGER, C. F. (1984). Children’s numerical estimation: Flexibility in the use of counting. *Journal of Educational Psychology*, **76**, 55-64.
- QUANDT, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**, 873-880.
- 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. A. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 81-106). Oxford: Clarendon.
- SIEGLER, R. S., & ROBINSON, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 241-312). New York: Academic Press.
- VERSCHAFFEL, L., DE CORTE, E., LAMOTE, C., & DHERT, N. (1998). The acquisition and use of an adaptive strategy for estimating numerosity. *European Journal of Psychology of Education*, **13**, 347-370.

APPENDIX

The computer program SegcurvN is written in Turbo Pascal and runs on an MS-DOS IBM-compatible computer. The program’s characteristics and options are largely the same as those of Segcurve, described by Beem (1995). Therefore, the reader should consult Beem (1995) or the program manual (Beem, 1999) for details. The main difference is that the user must specify the number of change points and whether they are joins or jumps. Several combinations of join and jump models can be estimated in one run of the program.

Testing Linear Hypotheses

About the Parameters

The program can fit the models subject to any linear restriction on the parameters that the user may wish to make. Since the program's output contains various measures of fit (among them the models' error sum of squares, or SSE), hypotheses can be investigated by comparing the measures of fit under various restrictions on the parameters. Alternatively, the hypotheses can be tested formally by computing significance tests. If $SSE(H_1)$ and $SSE(H_0)$ are the error sum of squares under hypotheses H_1 and H_0 , where H_0 is nested in H_1 (i.e., H_1 is the least, and H_0 the most, restrictive hypothesis, and H_0 is a special case of H_1), then the statistics are of the F type,

$$\frac{n[SSE(H_0) - SSE(H_1)]}{SSE(H_1)},$$

and the log likelihood ratio statistic:

$$n(\ln[SSE(H_0) \times SSE(H_1)]).$$

Here n is the number of observations. The statistics are both distributed asymptotically as χ^2 on q degrees of freedom, where q equals the number of linearly independent restrictions under H_0 minus the number of linearly independent restrictions under H_1 . The program can write all the computed statistics and parameters to a file that is readable by other programs. Therefore, the user can subsequently perform such analyses with other programs, such as spreadsheets.

Testing for Two Change Points

The program can be instructed to compute recursive residuals and cumulative sums (or cusums) of recursive residuals in order to test for a change point. Cusums form the basis for statistics that can be used to test for the constancy of a regression equa-

tion over a range of values of the independent variable x (Brown et al., 1975; Schweder, 1976). The program computes the different statistics, including the cusums and the boundary values, that are needed to test for a change point. If the cusums exceed these boundary values, the null hypothesis of no change point is rejected. The probability that the cusums exceed one of these bounds is equal to the predetermined α level. To guarantee a sufficiently large power, Beem (1993, 1999) suggested an α level of .10 for this test.

Showing that more than one change point occurs by more than one crossing of the boundaries of the cusums requires a large number of observations and many different values of the independent variable. The occurrence of two change points in a specific data pattern can, however, be tested with more power by dividing each subject's data pattern into two parts, but at the cost of conditioning on change points that have been estimated. The first part, then, contains all data points that were included in the first and second segments of the regression model, whereas the second part consists of all data points of the second and third segments. Now, both parts can be considered as separate two-phase models and thus they can both be tested for the presence of one change point. When the cusum test yields a change point in both models, the presence of a three-phase model is confirmed in that particular data pattern.

NOTE

1. However, because this statistical test was too stringent according to our definition of the different patterns— b_2 needed to have a value *close to zero*, not *equal to zero* in Patterns 2 and 4—all b_2 parameters that lay within the range of $[-200, 200]$ were also considered as evidence for the use of the estimation strategy.

(Manuscript received December 27, 2000;
revision accepted for publication June 2, 2001.)