Perception & Psychophysics
1978, Vol. 24 (2), 182-184

Extrapolation of exponential time series is not enhanced by having more data points

W. A. WAGENAAR
Institute for Perception TNO, Soesterberg, Holland

and

H. TIMMERS
Institute for Perception Research, Eindhoven, Holland

Subjective extrapolation of time series can formally be described as a two-stage process. In the first stage, some properties of the time series are identified; whether the series is stationary or not, whether there is an increase or decrease, acceleration or deceleration, whether or not there are periodicities, and so on. In the second stage, the rule previously discovered is applied in generating further elements of the series. Both stages are open to error; acceleration in a nonlinear series can be overlooked (Stage 1), or the subject may produce the third future element instead of the fifth (Stage 2). Whenever subjects fail to correctly extrapolate a time series, one would like to know whether either one or both stages contribute to the inaccuracy.

Previous experiments on the extrapolation of exponential processes showed that subjects tend to make mistakes in Stage 1 rather than in Stage 2. The evidence was that subjective extrapolations in a variety of conditions could adequately be described by a simple model which assumed that the subject applied a constant, but underestimated, growth factor. If the time series is described by:

$$y = e^{bx}, \tag{1}$$

where $x = 1, 2, \ldots, n$, the growth factor is e^b. Subjective continuation of the series was perfectly described when a growth factor of $\alpha e^{\beta b}$ was used; α was equal to 1.0 or a little above, β was typically below 0.2 (Wagenaar & Sagaria, 1975). Factors affecting the value of β were related to the mode of presentation: tables are better than graphs, ascending curves are worse than descending curves (Timmers & Wagenaar, 1977). The value of β was not affected by a variable expected to influence the second stage: year-by-year extrapolation vs. extrapolation in one large step. The year-by-year extrapolation showed that subjects use the same multiplier $\alpha e^{\beta b}$ in all successive steps; this was the most direct evidence that subjects use the correct rule but the wrong parameters.

The present study is designed to investigate another variable that may influence the first stage. This variable is the number of data points in the starting series. The paradigm is simple. In previous studies, subjects were always presented with five successive data points, e.g., 3, 7, 20, 55, 148 ($y = e^x$, $x = 1, 2 \ldots, 5$). The trajectory 3 through 148 can also be covered by three data points (3, 20, 148) or seven data points (3, 5, 10, 20, 39, 76, 148). These three representations describe the same exponential function, but with a different sampling frequency. The number series can be mathematically described by second-, fourth-, or sixth-order polynomials. Since such functions increase faster when more high-order terms are added, it is to be expected that subjects, fitting the simplest polynomial, would produce higher extrapolations when more data points are available. This is a prediction put forward by Jones (1977; see also Wagenaar, 1977). The example presented in Table 1 may serve to illustrate the point. In contrast, Timmers and Wagenaar (1977) suggested that subjects do no such thing as fitting a polynomial, but rather that they intuitively estimate a growth factor by inspecting a few initial differences between successive data points. This would explain why descending curves are better extrapolated. Obviously, differences between successive data points become smaller the more data points are presented. Accordingly, it may be expected that extrapolations based on these differences decrease when more data points are presented; in short, the more data points, the worse the extrapolations. This phenomenon has been observed before with nonnumerical presentations of exponential growth (Wagenaar & Timmers, 1978, in press). In the present experiment, the prediction was tested with numerically presented data points.

Table 1
Three Sets of Data Points with Extrapolations Based on the Simplest Polynomials Describing the Data Points

n	Data Points							Extrapolations According to the Best-Fitting Polynomial with Power n − 1					
3	3			20			148			387			737
5	3		7	20		55	148		358	767		1480	2625
7	3	5	10	20	39	7	148	283	523	928	1576	2569	4035

Note—n = number of data points. Exponential curves fitted to the data points result in an extrapolation of about 8100 on the last position.

Table 2
Survey of the Seven Experimental Conditions

	Number of Years Covered by the Data Points		
n	3	5	7
3	16	18	
5	19	18	21
7		20	16

Note—n = number of data points. The entries indicate the number of subjects within groups.

METHOD

Stimulus Materials

Starting functions were of the general form presented by Formula 1. In the case of five data points (i.e., $x = 1, 2, \ldots, 5$), the values of b employed were 1.0, 1.2, 1.4, and 1.6. Thus, the starting numbers were between 3 and 148, 3 and 403, 4 and 1,097, 5 and 2,981, respectively. The same trajectories were also described by three and seven data points, by taking, respectively, $x = 1, 3, 5$ and $x = 1, 1.7, 2.3, 3, 3.7, 4.3, 5$.

The total number of experimental functions was 3 (number of data points) $\times$ 4 (values of b). These functions were mixed with 36 other functions, obtained by adding a constant of about 100, 200, or 300 to the experimental functions. The aim of these "filler functions" was to provide a contextual range of functions similar to the one used in previous experiments.

Procedure

The written instruction explained that many processes in our environment, such as pollution, inflation, and energy consumption, show marked growth tendencies, and that it is sometimes important that individual citizens perceive these tendencies correctly. Then an example was presented with the phrase: "In the last (k = 3, 5, or 7) years, the following observations were obtained (then followed the data points): which value will be reached after another k years, provided that the process goes on uninhibited?"

The number of years mentioned was not always the same as the number of data points. The seven combinations used are presented in Table 2. One can conceive of the experiment as an aggregate of three subexperiments: one on the effect of the number of data points when the number of years covered by the data points is kept at 5; one on the effect of the number of years covered by the data points when the number of data points is kept at five; one on the effect of number of data points and number of years together, when they are kept equal. This setup asked for a between-subjects design with one group of subjects in each condition.

Within each group, subjects were presented with 16 extrapolation problems: 4 experimental problems and 12 filler problems. The order of presentation was varied systematically across subjects. The experiment was run as a classroom experiment. The time allotted to each problem was 20 sec.

Subjects

Seven groups of subjects were used in the conditions presented in Table 2, and the number of subjects was as indicated. The subjects were 128 undergraduate psychology students at the University of Leyden.

RESULTS AND DISCUSSION

The model based on extrapolations according to the correct rule but using the wrong parameters predicts the response to be:

$$\hat{y} = e^{5b} \cdot \alpha^{(n-1)} e^{4\beta b}, \tag{2}$$

in which e^{5b} is the last data point presented, n is the number of data points presented, and α and β are parameters to be estimated from the data. The willingness to produce large numbers, α, is raised to the power $n-1$ because it is assumed that the extrapolation is performed in $n-1$ steps. Alternative possibilities are discussed later on. Taking natural logarithms on both sides of Formula 2 yields:

$$\ln \hat{y} = b(5+4\beta)+(n-1)\ln \alpha. \tag{3}$$

Thus, if $\ln \hat{y}$ is plotted against b, we should obtain a linear function with intercept $(n-1)\ln\alpha$ and slope $(5+4\beta)$. Such curves are presented in Figure 1. The design for an ANOVA with unequal group size is presented in Winer (1962, p. 374).

Number of Data Points

There was a significant effect of the number of data points in the three conditions in which the starting numbers were supposed to cover 5 years [$F(2,53) = 3.22$, $p < .05$]. The difference between three and seven data points was about a factor of three, which is quite substantial. Extrapolation according to the best fitting polynomial could account for the results in the condition with three data points, but certainly not for the results in the two other conditions (see Figure 1).

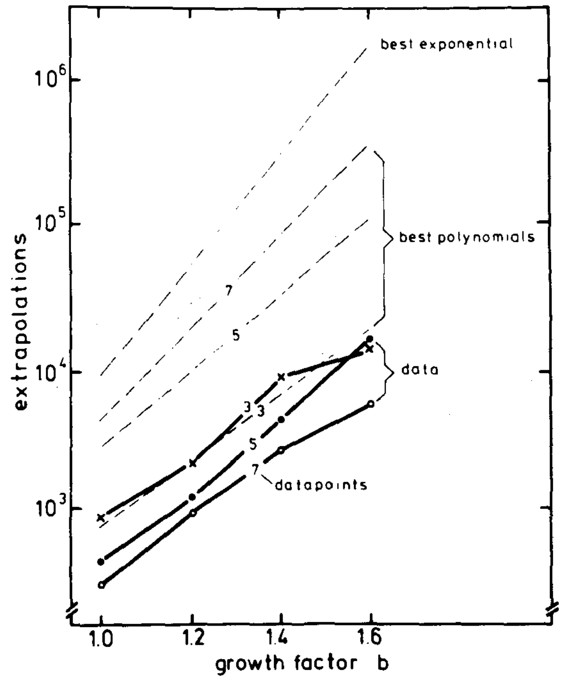

Figure 1.

Number of Years Covered by the Data Points

In the three conditions in which five data points were presented, we were able to test the effect of the number of years covered by the data points. This effect was negligibly small and not significant [$F(2,55) = 1.60$, n.s.].

Number of Data Points and Number of Years Varied Together

In three groups, the number of years and data points were kept equal. The differences among these three groups were in the same direction as the differences mentioned before, but not quite significant.

Number of Data Points Irrespective of Number of Years

Since number of years did not have any effect, it seemed reasonable to combine groups presented with the same numerical problems. The effect of number of data points again came out very significant [$F(2,125) = 5.40$, $p < .01$].

The results confirm our hypothesis that extrapolation of exponential functions improves when less data points are shown. Is this improvement due to an increased sensitivity to growth or to an overall willingness to produce large numbers? To answer this question, we fitted functions as described by Formula 3 to the data and obtained the best estimates of α and β. These values are presented in Table 3. It is clear from this table that β, the proportion of the exponent taken into account, is very small and not responsible for the better extrapolations in the conditions with less data points. Rather, the willingness to produce large numbers, α, is affected when many data points are presented. Covering a trajectory by a few data points gives a subject the general feeling that "it is going to be a large number" but it does not increase sensitivity to growth.

In the introduction, it was mentioned that number of data points was conceived of as a variable affecting the first stage of the extrapolation process, i.e., perception of the rule behind the data. In our Formula 3, however, it is assumed that subjects extrapolate step by step and that they take $(n-1)$ steps, n being the number of data points. Hence, number of data points seem to affect the second stage as well; the values of α in Table 3 are based on this assumption. This is not necessarily a complication, however: one can easily describe Stage 2 as taking one giant step from the last data point to the required response. This will change the estimates of α but not those of β. The values of α just become the aggregate values $\alpha^{(n-1)}$, amounting to 7.63, 2.23, and 1.87 for, respectively, three, five, and seven data points. When Stage 2 is described as an application of the discovered rule a fixed number of times m, m being independent of the number of data points presented, values of α can be obtained by taking the m^{th} root of the values presented above. This will not affect the clear-cut relation between α and the number of data points.

Table 3
Model Parameters Fitted to the Data of All Groups

N	Variance*	Intercept	Slope	α	β
3	.96	2.03	4.63	2.76	−.09
5	.99	.80	5.28	1.22	.07
7	.99	.63	5.13	1.11	.03

Note—N = number of data points.
**Proportion of variance explained by linear trend.*

REFERENCES

Jones, G. V. Polynomial perception of exponential growth. *Perception & Psychophysics*, 1977, **21**, 197-198.

Timmers, H., & Wagenaar, W. A. Inverse statistics and misperception of exponential growth. *Perception & Psychophysics*, 1977, **21**, 558-562.

Wagenaar, W. A. Polynomial perception of exponential growth: A reply to Jones. *Perception & Psychophysics*, 1977, **21**, 199.

Wagenaar, W. A., & Sagaria, S. D. Misperception of exponential growth. *Perception & Psychophysics*, 1975, **18**, 422-426.

Wagenaar, W. A., & Timmers, H. Intuitive prediction of growth. In D. F. Burkhardt & W. N. Ittelson (Eds.), *Environmental assessment of socioeconomic systems*. London: Plenum, 1978. Pp. 103-122.

Wagenaar, W. A., & Timmers, H. The pond-and-duckweed problem: Three experiments on the misperception of exponential growth. *Acta Psychologica*, in press.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received for publication May 9, 1978; accepted May 24, 1978.)