DOCUMENT RESUME

ED 306 041 PS 017 993

AUTHOR	Burchinal, Margaret R.
TITLE	Comparison of Models for Estimating Individual Growth Curves.
PUB DATE	Apr 89
NOTE	10p.; Paper presented at the Biennial Meeting of the Society for Research in Child Development (Kansas City, MO, April 27-30, 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Estimation (Mathematics); *Goodness of Fit; Guidelines; *Individual Development; *Mathematical Models; *Research Design; Simulation

ABSTRACT

Growth curve models are a useful tool for developmentalists because they can estimate an attribute's developmental function by providing a mathematical description of growth on an attribute over time. However, selection of a growth curve model appropriate for estimating individual developmental functions is problematic. The ideal model is the one that most precisely estimates individual developmental functions from the profile data. But profile data often violate model assumptions. When sample sizes are relatively small, the effects of these violations often are not well understood. For this study, computer simulations were run to identify which types of models provided the most precise descriptions of developmental functions with various types of profile data. Models included: (1) Population Logistic Growth Curve; (2) Population Polynomial Growth Curve; (3) Individual Logistic Growth Curve; (4) Individual Polynomial Growth Curve; and (5) Prototypic Growth Curve. The goals of three analyses were to identify the best model for estimating growth curves when individual differences and reliability are varied, when profile size and reliability are varied, and when more than one parametric family is sampled. All examined data characteristics affected the ability of the models to estimate the profiles. It is concluded that longitudinal studies must be carefully designed if data are to be used to estimate individual growth curves. (RH)

Comparison of Models for Estimating Individual Growth Curves

Margaret R. Burchinal

Frank Porter Graham Child Development Center
Bypass 54 West (072A) CB #8180
University of North Carolina at Chapel Hill
Chapel Hill, N.C. 27599-8180

Paper presented at the Biennial meeting of the Society for Research in Child Development, April 1989, Kansas City.

Comparison of Models for Estimating Individual Growth Curves

Growth curve models are a useful tool for developmentalists because they can estimate an attribute's developmental function by providing a mathematical description of growth on an attribute over time. However, selecting an appropriate growth curve model to estimate individual developmental functions is problematic; the ideal model is the one that most precisely estimates individual developmental functions from the profile data, but profile data often violate model assumptions. The effects of these violations may be known asymptotically, but often are not well understood with relatively small sample sizes. Accordingly, computer simulations were run to identify which types of models provide the most precise descriptions of developmental functions with various types of profile data. These simulations will be discussed after a few introductory comments.

The type of growth curve methods that I examined describe change across time on one attribute that had been measured with an interval or ratio level scale. The ability to measure the attribute and the appropriateness of the model limit the ability to estimate individual growth curves. The "true" developmental function can be estimated only when the attribute has been measured isomorphically and the appropriate growth curve function has been selected. However, too often in child development we can not measure attributes of interest isomorphically (i.e., scores indicate how much of the attribute the individual has at that time)-- instead we use relativistic measurement (e.g., scores indicate the relative ranking of the individual within some normative population). While the "true" developmental function can not be estimated with relativistic scores, individual patterns of change can still be estimated when appropriate growth curve models are selected.

Selection of the growth curve model also depends on the investigator's assumptions and knowledge about growth of the attribute. I classified methods along 3 dimensions: type of function, for whom the function is estimated, and type of estimation. Developmentalists have estimated linear and nonlinear functions, but nonlinear functions are usually necessary to describe the growth periods. "Inherently" nonlinear functions usually provide the most information about growth because they are based on assumptions about the growth process and have intrinsically meaningful parameters such as the asymptotic level, rate of change, and time at which half of the growth has occurred. Polynomial growth curve (PGC) models such as the quadratic or higher-order curves can provide good approximations of the inherently nonlinear functions, at least during the growth period. However, an appropriate "inherently" nonlinear function such as the logistic or exponential curve provides more information about the growth process.

These PGC and inherently nonlinear functions can be estimated to describe growth within an entire population or an individual. Population models are growth curve models that estimate a single curve for a given population. They assume that individual differences are trivial. Individual models estimate separate growth curves for each individual. They assume that individuals vary and that each individual may show a different pattern of change over time.

These models can be estimated with ordinary least squares, weighted least squares, or maximum likelihood methods.

In addition, prototypic growth curve methods can identify individuals with similar patterns of change without requiring specification of a function. Cluster, p-type factor, or principal components analysis can be used to identify prototypic patterns of change within a sample. In this case, it is assumed that a relatively small number of patterns will characterize development for the sample of interest. This approach provides less information about the growth process; only predicted values are estimated, not the growth curve and its parameters.

I examined the "best-case" situation; the simulations corresponds to the study in which the attribute was measured isomorphically with random error. The most appropriate growth curve model, a good PGC approximation, and the prototypic growth curve model were estimated from the data.

In the simulations, profile data on hypothetical individuals with known developmental functions were created. Within a sample either all of the individual developmental functions were logistic growth curves

$$Y_{ij} = d_i \,/\, \{1 + \exp[-g_i\,(T_{ij} - a_i)]\},$$

or half of them were logistic and half were exponential,

$$Y_{ij} = d_i * \{1 - \exp[-g_i\,(T_{ij} - a_i)]\}.$$

Figure 1 displays the expected growth curves from the simulations. The individual differences among developmental functions were either nonexistent or small (i.e., the parameters of the individual functions were sampled from normal distributions with small variances). The individual's error-free profile was created by observing the values that the developmental function assumed at either 5 or 10 time points. Independent random error was added to the error-free profiles such that reliability was either very high (.95) or moderate (.80). Thirty individual developmental functions were generated for each of 1000 replications of seven independent cells.

The effects of varying both data and model characteristics on the relative ability of selected growth curve models to estimate the developmental functions from the errorful profile data were examined. Table 1 lists the selected growth curve models. The population logistic growth curve was estimated from the sample's data with ML, using iterative OLS estimates (Gallant, 1975a) and using SUNR (Gallant 1975b). The population cubic curve was estimated from the entire sample's data using the multivariate approach and orthogonal polynomial contrasts, using OLS (Potthoff & Roy, 1964) and WLS (i.e., the higher order contrasts were included as covariates, Rao (1966)). The individual logistic curves were estimated from each individual's data separately with ML (iterative OLS) only. The individual cubic curves were estimated first by OLS (i.e., fitting the cubic model to each individual's data separately with OLS) and using the Empirical Bayes Mixed Model approach (Fearn, 1975; Laird & Ware, 1982). Finally, prototypic growth curves within the sample were estimated with a truncated principal components analysis (Overall & Klett, 1972; Nunnally, 1962; 1978).

Three analyses examined the effects of manipulating various data characteristics on the ability to estimate developmental functions (see Table 4). The primary criterion was the mean squared distance between the "true" and estimated growth curve. The first analysis asked which model "best" estimated the individual growth curves when the degree of individual differences and reliability of measurement were varied. The second analysis identified which model best estimated the growth curves when profile size and reliability were manipulated. The final analysis examined the effects of mixing developmental functions.

In general, the results indicated that the growth curve model that most closely approximates the developmental functions and whose assumptions are least likely to be violated by the data tended to provide the best fit to those data. Table 3 displays the results form analyses 1-3. Analysis 1 indicated that models that estimated a population growth curve were dramatically effected by whether nontrivial individual differences existed, but not by the reliability of the data. In contrast models that estimated separate curves for each individual provided much better estimates when data were highly reliable, but not by whether individuals differed. Only the prototypic and EB cubic curve models were effected by both.

Analysis 2 suggested that the individual growth curve models provided better estimation when data were highly reliable or when many observations per individual were collected. Precision was markedly poorer when these models were fit to less reliable data consisting of 5 observations. This trend was also observed with the PCA approach, but not with most of the population models.

Analysis 3 demonstrated that even seemingly minor violations of certain assumptions resulted in very poor fits. Using the parametric family of the developmental functions (LGC) was clearly preferable to using an approximation, but only when the data are all from that parametric family.

Table 2 displays the mean squared distance between error-free and predicted profiles. The columns of this table correspond to the 7 types of data examined while the rows list mean squared distances for the selected growth curve models. The boxes encompass the types of data in which that a given model provided the among the best estimation of the error-free profile data. Individual developmental functions tend to be estimated with relative precision when the selected model is fit only to the individual's data (individual PGC or LGC) if the developmental functions vary nontrivially and to the entire sample's data (population LGC or PGC) if individual differences are trivial. In addition, the prototypic growth curve model tends to recover the error-free scores well from all types of data examined in this study, but provide less information about the growth process. Finally, comparisons of OLS and approximate WLS methods suggest that OLS uniformly provides more precise estimates when individuals vary nontrivially. These findings suggest that the number of observations in each profile and the reliability of measurement will interact with both the type of growth curve model used for analysis and whether individuals' error-free profiles differ significantly across time.

These results imply that careful design of longitudinal studies is necessary if the data are to be used to estimate individual growth curves. Factors that should be considered during the design stage include selection of growth function (i.e., ideally these should an appropriate nonlinear function if the attribute was measured isomorphically and a PGC if scores do not represent the attribute isomorphically. The investigator should determine whether individual differences are nontrivial. If they are, then reliability should be high and/or profile size should be large. This study suggests precise estimation of individual growth curves can occur only when the design factors such as reliability and profile size were considered before data were collected.

In conclusion, the simulations indicate that one should select the growth curve model whose assumptions are least likely to be violated by the profile data and whose function mostly closely approximates the "true" individual growth curve. However, seemingly minor violations of model assumptions can produce poor fits. All examined data characteristics (number of parametric families, presence of individual differences, profile size, and reliability) effected the ability of the selected growth curve models to estimate the error-free profiles. The amount of information provided about development was inversely related to the variety of conditions that goodness of fit criteria were met. Finally, recent work in linear and nonlinear mixed models looks very promising in terms of increasing the precision of estimation for individual growth curves, even in the presence of missing or mistimed observations, when individual curves are normally distributed about the population curve.

REFERENCES

Fearn, T., (1975). A Bayesian approach to growth curves, *Biometrika*, *62*, 89-100.

Gallant, A. R. (1975a). Nonlinear regression. *American Statistician*, *29*, 73-81.

Gallant, A. R. (1975b). Seeming unrelated nonlinear regression. *Journal of Econometrics*, *3*, 35-50.

Laird and Ware (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.

Nunnally, J. C. (1962). The analysis of profile data. *Psychological Bulletin*, *59*, 311-319.

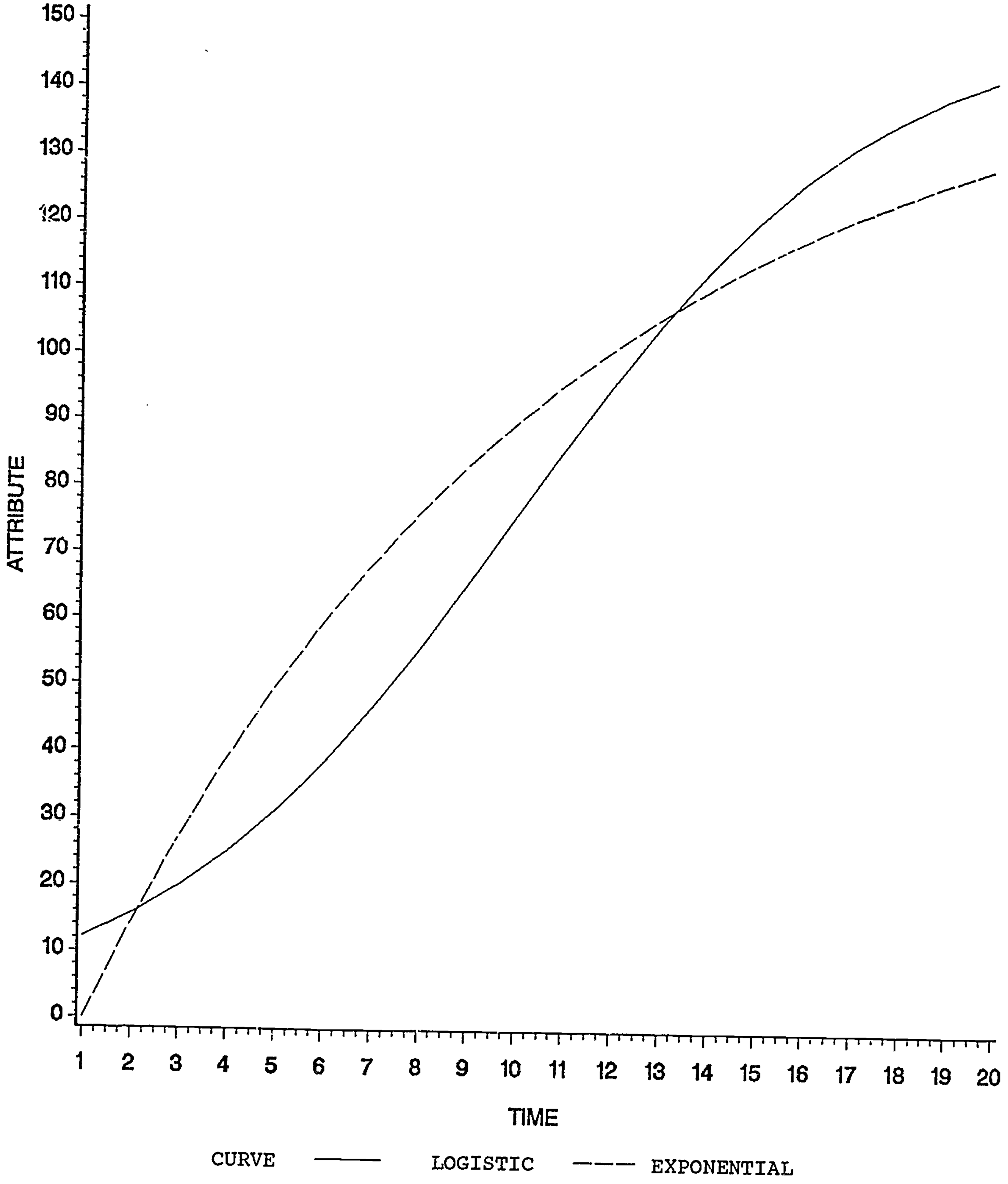
Nunnally, J. C. (1978). *Psychometric Theory*, New York: McGraw Hill Publishing Co.

Overall and Klett (1972). *Applied Multivariate Analysis*, New York: McGraw Hill Publishing Co.

Potthoff and Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

Rao, C. R., (1966). Covariance adjustment and related problems in multivariate analysis. In P.R. Krishnaiah (Ed.) *Multivariate Analysis*. New York: Academic Press, 87-103.

EXPECTED LOGISTIC AND EXPONENTIAL GROWTH CURVES



LOGISTIC: $E(Y_i)=151 / (1 + \exp(-.27(T_i-10)))$
EXPONENTIAL: $E(Y_i)=151 * (1 - \exp(-.10(T_i -1)))$

Table 1

Selected Growth Curve Models

Population Logistic Growth Curve[a]	$E(Y_{ij}) = d / (1 + \exp(-g * (T_j - a)))$,
Population Polynomial Growth Curve[b]	$E(Y_{ij}) = \beta_0 + \beta_1 * T_j + \beta_2 * T_j^2 + \beta_3 T_j^3$
Individual Logistic Growth Curve[a]	$E(Y_{ij}) = d_i / (1 + \exp(-g_i * (T_{ij} - a_i)))$,
Individual Polynomial Growth Curve[b]	$E(Y_{ij}) = \beta_{0i} + \beta_{1i} * T_{ij} + \beta_{2i} * T_{ij}^2 + \beta_{3i} T_{ij}^3$
Prototypic Growth Curve[c]	$E(Y_{ij}) = P_{1i} F_{1j} + P_{2i} F_{2j} + \dots + P_{ri} F_{rj}$

Note: [a] d is the asymptotic level, g is a rate of change parameter, and a is the age at which half of the growth has occurred

[b] the 3 parameters are the intercept, linear slope for age, and quadratic slope respectively

[c] $\underline{P}$ is a weight matrix and $\underline{F}$ is a matrix whose columns represent the prototypic growth curves.

Table 2
Manipulated factors in three analyses.

	Manipulated Factors			
	Parametric Family	Individual Differences	Reliability of data	Profile Size
Analysis 1	logistic	none or small	95% or 80%	5
	Goal: Identify the "best" model for estimating growth curves when individual differences and reliability are varied			
Analysis 2	logistic	small	95% or 80%	5 or 10
	Goal: Identify the "best" model for estimating growth curves when profile size and reliability are varied			
Analysis 3	logistic or logistic and exponential	small	95%	5
	Goal: Identify the "best" model for estimating growth curves when more than one parametric family is sampled			

Table 3
Effects of manipulated data characteristics.

Analysis	Growth Curve Model							
	Population				Individual			Prototypic
	LGC		PGC		LGC	PGC		PCA
	OLS	WLS	OLS	WLS	OLS	OLS	WLS	OLS
1: Effects of Reliability and Individual Differences								
Reliability	ns	ns	ns	ns	***	***	***	**
Individual Differences	***	***	***	***	ns	ns	**	ns
Reliability * Ind. Dif.	ns	ns	ns	ns	ns	ns	*	*
2: Effects of Profile Size and Reliability								
Profile Size	ns	ns	ns	**	**	***	**	*
Reliability	ns	ns	ns	*	***	***	***	**
Profile Size * Reliability	ns	ns	ns	*	*	**	**	ns
3: Effects of Sampling More than One Parametric Family								
Parametric Family	xxx	xxx	xxx	xxx	xxx	ns	x	x

Note: * p(F(1,3996)) < .0001, ** F(1,3996) >1,000, *** F(1,3996) >10,000

x p(F(1,1996)) < .0001, xx F(1,1996) >1,000, xxx F(1,1996) >10,000

Table 4

Squared Distance between Predicted and Error-free Profiles

Growth Curve Models		Manipulated Factors						
		Type of Developmental Function						
		LGC						LGC & EGC
		Individual Differences						Ind. Dif.
		None		Small				Small
		Profile size		Profile Size				Profile Size
		5		5		10		5
		Reliability		Reliability		Reliability		Reliability
		.80	.95	.80	.95	.80	.95	.95
Population LGC OLS	m	0.07	0.02	13.85	13.72	13.57	13.67	60.79
	std	0.06	0.01	3.22	3.29	3.27	3.29	5.22
Population LGC WLS	m	0.07	0.02	13.88	13.74	13.65	13.74	68.86
	std	0.07	0.01	3.23	3.30	3.29	3.31	6.10
Population PGC OLS	m	0.10	0.02	13.96	13.82	14.15	14.25	54.63
	std	0.07	0.02	3.25	3.32	3.28	3.29	5.16
Population PGC WLS	m	0.10	0.02	14.40^{a}	14.30	17.34	25.57	56.28
	std	0.08	0.02	3.35	3.43	4.30	8.85	5.34
Individual LGC OLS	m	5.26	1.12	5.26	1.10	1.49	0.32	25.55
	std	0.78	0.16	1.45	0.30	0.41	0.09	0.72
Individual PGC OLS	m	14.04	3.00	14.20^{a}	3.10	3.30	1.48	3.08
	std	1.81	0.38	3.73	0.81	0.65	0.17	0.87
Individual PGC WLS	m	0.49	0.11	28.80	9.55	7.85	3.95	10.85
	std	0.35	0.08	8.40	3.00	1.45	0.47	4.23
PCA	m	4.27	0.94	3.32	1.54	2.23	0.61	1.79
	std	1.80	0.38	1.29	1.94	0.80	0.13	0.44

note: superscripted letters indicate means that did not significantly differ ($p<.0001$) in pairwise comparisons within condition. That is, only the Population PGC WLS and Individual PGC OLS models when fit to moderately reliable logistic data when $p=5$ and individual differences were small did not differ significantly.