USING FACTOR ANALYSIS TO ESTIMATE PARAMETERS

BY

JAGDISH N. SHETH

Reprinted from the Journal of the American Statistical Association
September 1969, Vol. 64
pp. 808-822

USING FACTOR ANALYSIS TO ESTIMATE PARAMETERS

Jagdish N. Sheth
University of Illinois

Factor analysis has been traditionally utilized for three broad purposes: First, as a data reduction technique which will hopefully simplify a multivariate situation to a smaller set of dimensions and enable the researcher to utilize data on a large number of variables; second, as an indexing device in which overtly manifested data are transformed to provide the latent or unobservable trait of a phenomenon; finally, as a cluster technique which helps the researcher to classify a variety of observations into a small set of clusters which are also generally ordered. All three approaches have used a common starting point, namely, a matrix of correlations either among the variables (R type factor analysis) or among the observations (Q type factor analysis). In other words, factor analysis has been limited to associative relations.

The same technique can be used in *functional* relations, especially to estimate parameters of all linear and several nonlinear functions. It, therefore, essentially seems to serve the same function as curve fitting techniques but with two important distinctions. First, the procedure outlined here provides estimates of parameters for *each observation* (individual) in the sample as well as aggregate parameters. Second, it provides decision rules for various types of functions where aggregation of observations (individuals) assumed to have certain functional relations is or is not legitimate.

1. ESTIMATING PARAMETERS OF FUNCTIONAL RELATIONS

Rao [4] and Tucker [7] both independently suggested the possibility of using factor analysis in measuring growth or learning curves. Relatively little is known about their suggestions in many applied disciplines despite the fact that there are several areas of investigation where their theoretical contribution seems potentially useful. More importantly, one may easily extend their suggestion to functional relations other than the special cases of exponential functions like growth or learning curves.

In general, we may state that factor analysis can be used in all instances where the independent variable is some function of time and the dependent variable is observed at discrete time intervals. The model is then relevant for several areas of management, for example, measuring trends in sales forecasting, or development of brand loyalty of consumers over time, or rate of defects in a product as a function of time. Similarly, it can be used to calculate accident proneness of factory workers, deal proneness of consumers, and absence proneness of clerical staff, to mention a few. In most of these areas, the common approach has been to use a variety of stochastic processes.

Take a function

$$y = f(a, b, c, \cdots, x) \tag{1}$$

where y represents the dependent variable and x represents the independent time variable, and a, b, c, etc. are constants of the function. In consumer behavior, for example, y may be number of units of a brand purchased by the

consumer and x may be equal-interval time periods or trials; a, b, c, etc. are then parameters of the function, namely, the magnitude of purchase over time.

If aggregate parameters were desired, one could use the curve fitting methods. However, our interest is in estimating parameters of each individual's function in the sample. The individual functions can be represented by

$$y_1 = f(a_1, b_1, c_1, \cdots, x)$$
$$y_2 = f(a_2, b_2, c_2, \cdots, x)$$

or more generally

$$y_i = f(a_i, b_i, c_i, \cdots, x). \tag{2}$$

Furthermore, the dependent variable is measured at discrete points in time, say x_j. Then

$$y_{ji} = f(a_i, b_i, c_i, \cdots, x_j). \tag{3}$$

Thus an observation y_{ji} is a function of parameters of the individual i and the value of the independent variable, x_j. Several functions, both linear and nonlinear, can be transformed[1] to produce

$$y_{ji} = \sum_{m=1}^{r} f_m(x_j)F_m(G_i), \qquad G_i = a_i, b_i, c_i, \text{etc.} \tag{4}$$

The $f_m(x_j)$ are functions of the independent variable x_j. The $F_m(G_i)$ are similarly functions of the parameters a_i, b_i, c_i, etc.

Depending on the type of individual functional relations, the number r in equation (4) may be finite or infinite. In the latter case, equation (4) represents an infinite series, and very often, a small number of terms of the series will yield an adequate approximation to y_{ji} [7].

An observation y_{ji} which is a result of the interaction of an individual i with some environmental influence at time x_j, is, therefore, considered as a sum of the products of functions of the individual parameters and corresponding functions of the independent variable. By the transformation process in equation (4), we obtain *derived* parameters which have the properties of determining the particular function for each individual.[2] This transformation process analyzes a functional relation into linear components, $m = 1, 2, \cdots r$. In other words, it treats a functional relation as being *multidimensional* and each component is one dimension [8]. To facilitate communication, each linear component will be called a 'reference curve', following Tucker [8]. A functional relation may have a *family* of reference curves ($1 \leq m \leq r$) depending upon its complexity.

The problem now is to find some technique which will generate $f_m(x_j)$ and $F_m(G_i)$. Equation (4) is analogous to the basic linear postulate of factor analy-

[1] The process of transformation into a linear system is similar to what is done with the raw data in curve fitting. For a good discussion see Tucker (7) and Merrill (3).

[2] The parameters are themselves functions of parameters. As functions, they lose the standard meanings of slope and intercept of primitive parameters. However, they do possess the relevant information related to shape of the curve so that it could be fitted to the data for an individual in the least squares sense.

sis. This can be easily seen if we define

$$f_m(x_j) \equiv a_{jm} \quad \text{and}$$
$$F_m(G_i) \equiv s_{mi}.$$

Then the basic factor analytic equation is obtained:

$$y_{ji} = \sum_{m=1}^{r} a_{jm}s_{mi}. \tag{5}$$

In the usual terminology, a_{jm} represents the 'factor loading' of variable j on factor m. Similarly, s_{mi} represents the 'factor score' of individual i on factor m. Finally, the observed score y_{ji} represents the weighted sums of the factor scores (s_{mi}, $m=1, 2, \cdots r$) for individual i and the weights are the factor loadings of the variables. However, when we use factor analysis for estimating parameters, the three components in the analysis acquire new meanings. Each reference curve (m) contains a series of a_{jm} parameter values which define it at all the discrete measurement points j of the independent variable. Each reference curve also reflects a component of the aggregate function. This will be elaborated later in the paper. Each s_{mi} value is the derived parameter estimate of an individual i's functional relation, and it performs the function of a weight to differentiate that individual from the function of the aggregate sample.

A brief description of factor analysis will be presented later. Before that, legitimacy of the transformation process for a variety of functions will be described.

2. TYPES OF FUNCTIONS WHERE TRANSFORMATION IS POSSIBLE

Not all functional relations can be transformed into equation (4). First, the resolution of a function into linear components entails expansion of the function by series analysis. There are several types of nonlinear functions such as parabolic and hyperbolic, which when expanded do not resolve into finite components. To that extent, they cannot be represented by a finite set of reference curves. However, in some cases, a finite number may adequately approximate the function, and so the transformation equation may be used. Second, and more important, factor analysis entails aggregation of observations on sample individuals as is true in most statistical analysis. Such aggregation becomes a serious problem when individual functions are aggregated. The aggregate function is modified by the averaging procedure in some individual functions. The transformation process then is not appropriate when individual functions are modified by averaging.

Estes [2] provides an elegant discussion of functions which are not modified by averaging. He classifies functions into three types: Type A, Type B and Type C. A formal criterion, suggested by Merrill [3], for classifying a function as Type A function is: If in the Taylor's series expansion of a function (rewritten such that each individual's parameters are treated as deviations from the mean values of the parameters) all second and higher order partial derivatives of the function with respect to the parameters are zero, it is a Type A function, and it remains unmodified as to form by averaging.

Class B functions are those for which a function $y=f(a, b, c, \cdots, x)$ does not satisfy the criterion mentioned above when expanded around $\bar{a}$, $\bar{b}$, $\bar{c}$, etc. but does satisfy that criterion when the function is rewritten as $y=f(u, v, w, \cdots, x)$ and expanded around $\bar{u}$, $\bar{v}$, $\bar{w}$, etc.; u, v, w, etc., are themselves now functions of the original parameters, a, b, c, etc.

Type C functions are those which when expanded by Taylor's series leave some second or higher order partial derivatives with respect to parameters of the function regardless of how the parameters are redefined.

All functions, linear and nonlinear, which are Type A or Type B functions can be transformed into equation (4) and can be resolved into exact finite components ($m=1, 2, \cdots r$). However, Type C functions will result in infinite components and could only be approximated by some small number of r components.

Some Examples

We take a few examples of each type of function to show how transformation is possible.

(i) *Type A Functions*

$$y = a + bx.$$

Then the observed value of an individual i at point j on x will be

$$y_{ji} = a_i + b_ix_j.$$

If we define

$$a_{j1} = 1 \qquad a_{j2} = x_j$$
$$s_{1i} = a_i \qquad s_{2i} = b_i$$

then

$$y_{ji} = a_{j1}s_{1i} + a_{j2}s_{2i}.$$

Therefore,

$$y_{ji} = \sum_{m=1}^{2} a_{jm}s_{mi}.$$

Thus for a simple linear equation possessing an intercept and a slope, we need a family of two reference curves, one of them a constant. Similarly, we can transform other Type A functions such as $y=a+bx+cx^2$ and $y=a(\log x)$.

(ii) *Type B Functions*

Here a function has to be rewritten so that we have some sort of transformation of the constants. For example, take the function

$$y = e^{b+x}.$$

Expanded as is, it will result in some second or higher order partial derivatives with respect to the parameter. If we redefine the function, however, as

$$y = (e^b)(e^x)$$

then

$$y_{ji} = (e^{b_i})(e^{x_j})$$

and

$$a_{j1} = e^{x_j} \qquad s_{1i} = e^{b_i}.$$

Therefore,

$$y_{ji} = a_{j1}s_{1i}.$$

Only one reference curve is needed and the function is unidimensional. Similar functions of Type B are, for example, $y=c+e^{b+x}$ and $y=\log(bx)$.

(iii) *Type C Functions*

Parabolic and hyperbolic functions of the form $y=ax^b$ and $y=ax^{-b}$ fall in this category. Also, exponential functions of the form $y=ae^{bx}$ belong to this category. They cannot be transformed into the linear factor equation without creating an infinite number of components. There are, however, two possible ways we may still use the method. First, a small number of components may adequately approximate the function, as mentioned earlier. Second, if we transform the *raw measures*, we may use the method on the transformed data. For example, take

$$y = ax^b.$$

Analogous to regression curve fitting we may give the function a linear expression by taking logarithms of both sides. Then

$$\log y = \log a + b(\log x)$$

which can be rewritten as

$$y'_{ji} = \log y_{ji} = \log a_i + b_i(\log x_j)$$

for an i individual's observed value at point j on x. Then

$$a_{j1} = 1 \qquad a_{j2} = \log x_j$$
$$s_{1i} = \log a_i \qquad s_{2i} = b_i$$

and

$$y'_{ji} = a_{j1}s_{1i} + a_{j2}s_{2i}.$$

However, we have to remember that raw data are transformed and therefore interpretation would be different.

When several individuals, each having a functional relation, are aggregated, as is necessary for factor analysis, the aggregation process makes the first reference curve approximate the mean curve. If there is only one reference curve, it is identical to the mean curve. However, if there is more than one reference curve, the second, third, etc. reference curves act as "correction terms" showing the deviations from the average tendency. They are similar to higher-order partial derivatives in Taylor's series expansion. This is important to remember because when we aggregate individuals with different forms of functional relations as would be very likely in any empirical situation, the interpretation of the s_{mi} coefficients for each individual become slightly more complex. If an individual, for example, has close to zero value of s_{1i} but a high value for s_{2i}, then he can be labelled as belonging to the second reference curve, and his observed raw scores would have the same shape as the shape of the second reference curve. This will become more meaningful later when we describe the empirical analysis.

3. DETERMINATION OF REFERENCE CURVES

We will briefly describe the procedure of determining the reference curves of a function to estimate parameters. Suppose we have a data matrix Y consisting of n rows $(j=1, 2, \cdots, n)$ and N columns $(i=1, 2, 3, \cdots, N)$. The n rows are the various j points of the independent variable x and the N columns are the individuals i in the sample. The cell value y_{ji} is, therefore, the observation of behavior of individual i at point j on the independent axis. In general, the rows of the data matrix are less than the columns $(n<N)$ making it a rectangular matrix.

Any complete $n\times N$ rectangular matrix Y can be resolved [1, 9] as

$$Y = U\Gamma W \tag{6}$$

where U is an $n\times n$ orthogonal matrix ($U'=U^{-1}$ and $UU'=I$), W is an $N\times N$ orthogonal matrix ($W'=W^{-1}$, and $WW'=I$) and Γ is an $n\times N$ diagonal matrix containing the λ_m principal roots as diagonal entries in the upper left section and zeros elsewhere. Columns of U are called left principal vectors of Y and the rows of W are called the right principal vectors of Y [8].

A matrix Y_r of rank $r(r<n)$ can be constructed utilizing only some of the roots and vectors which will give the best approximation of the data matrix Y in the least squares sense [1, 9]. It is constructed by taking the first r left principal vectors of U, the principal roots in Γ, and right principal vectors in W. In other words,

$$Y_r = U_r\Gamma_rW_r. \tag{7}$$

U_r is, therefore, an $n\times r$ section of U, Γ_r is an $r\times r$ section of the diagonal matrix Γ, and W_r is an $r\times N$ section of W.

If we let $U_r\Gamma_r=A$ and $W_r=S$, where A is an $n\times r$ matrix containing a_{jm} elements and S is an $r\times N$ matrix containing s_{mi} elements, we have

$$Y_r = AS = \left(\sum_{m=1}^{r} a_{jm}s_{mi}\right). \tag{8}$$

The matrix A then provides the estimates of the environmental parameters and the matrix S provides the individual parameters as discussed before. In factor analysis, matrix A is equivalent to the *factor loadings* matrix and matrix S is equivalent to *factor scores* matrix. In our case, each column $m(m=1$ to $r)$, with the series of a_{jm} running over the j consecutive points of the independent variable, becomes a reference curve. Each reference curve is one dimension of the aggregate functional relation in the data. There will be then as many columns m in the A matrix as there are reference curves.

The resolution of a data matrix Y into the product of three matrices U, Γ, and W is easily done if the matrix Y is post-multiplied by its transpose Y'. We obtain

$$YY' = (U\Gamma W)(W'\Gamma U') = U\Gamma^2 U'. \tag{9}$$

The matrix YY' contains the sums of squares and sums of cross products of elements of Y. We will call it a cross-products matrix to distinguish it from a covariance or a correlation matrix. It is, however, a square symmetric matrix like the other two matrices which enables us to calculate the *characteristic* roots and vectors. In fact, U contains the characteristic vectors of YY' and Γ^2 contains the characteristic roots of YY'. The characteristic roots are, therefore, *squares* of the principal roots of Y as defined earlier. Each characteristic root is ordered in terms of the magnitude of variance it explains in the total system. Taking the first set of r characteristic roots which together explain a certain amount of total variance, we obtain

$$A = U_r\Gamma_r. \tag{10}$$

Then $S = W_r$ can be easily obtained by

$$S = \Gamma_r^{-1} U'Y. \tag{11}$$

Finally, the approximate matrix Y_r is obtained as

$$Y_r = AS = U_r\Gamma_r W_r. \tag{12}$$

The rank of the original matrix Y is reduced to r from its order n. The rank r determines the complexity of the functions aggregated in the data matrix. Finally, the approixmate matrix is a least-squares solution [1, 9].

4. STANDARDIZATION OF DIFFERENT SAMPLE SIZES

The standard factor analytic procedures start with a correlation matrix. In deriving the correlation coefficients, the data are standardized by setting the mean equal to zero and the variance equal to unity. However, we lose two vital bits of information in the process, namely, the level and the scatter of original observations. Both are crucial in a functional analysis [5]. This is the reason why we factor analyze a cross-products matrix. It retains both types of information.

However, the analysis described earlier will produce estimates of the s_{mi} coefficients in the S matrix that are scaled so that we maintain $SS' = I$. Since the S matrix is a matrix of order $r \times N$, the resulting coefficients are also a function of the number N of individuals in the sample. Then, even if two sets of data differed only as to the sample size, the resulting coefficients would not be comparable. We must, therefore, rescale the S matrix into a matrix V so that the coefficients in V are independent of sample size. This is possible if we pre-multiply the S matrix by a scalar matrix $N^{1/2}$ whose diagonal elements are the square-roots of the sample size [6]. Thus

$$V = N^{1/2}S. \tag{13}$$

In order to maintain the basic relation among the three product matrices, we must rescale $U_r\Gamma_r = A$ to P so that

$$P = (U_r\Gamma_r)N^{-1/2} = AN^{-1/2}. \tag{14}$$

Then

$$Y_r = PV = AN^{-1/2}N^{1/2}S. \tag{15}$$

The matrix P contains the rescaled reference curves and the matrix S contains the rescaled individual parameters. Each column of P is, therefore, a component of the aggregate function and the first will approximate the mean curve.

5. TRANSFORMATION OF FACTOR ANALYTIC RESULTS

If our approach were strictly inductive and our primary interest were to fit a curve to all individual functions such that we obtained the best fit, the procedures outlined above would suffice. However, we use least squares principles of curve fitting also in situations where we have had a hypothesis or theory, and where the data were deductively collected from the real world. In such theoretically-determined functional relationships, it is always desirable to extend the analysis by one more step. This entails transformation of factor analytic results summarized in P and V matrices. It is analogous to rotation of axes commonly performed in factor analysis by imposing rules of simple structure or their variants in the hope that the factors may become meaningful.

However, the rules that govern transformation in the case of estimating parameters need not follow the traditional rules of simple structure. In fact, Tucker [8] shows that in obtaining learning curves for a variety of reinforcement schedules, the varimax rotation is inappropriate since it would attempt to put as many zeros as possible in the learning curves, which goes contrary to the theory. In some cases, one of the variants of simple structure may be meaningful, however.

Furthermore, it is possible to focus on transformation of either the V matrix (individual parameters) or the P matrix (aggregate parameters), although in factor analysis the focus is generally on the transformation of only the factor loadings matrix.

There are, however, no standard methods for transformation of individual and aggregate parameters because each situation may be uniquely specified by some theory or hypothesis. Therefore, a general procedure of transformation is described here.

Let T be a square, non-singular matrix of order r and define

$$Z = TV$$
$$Q = PT^{-1}$$

if we are interested in transforming the individual parameters. Then

$$QZ = PT^{-1}TV = PV = Y_r. \tag{16}$$

Conversely, we may be interested in meaningful representation of aggregate parameters so that

$$Q = PT$$
$$Z = T^{-1}V.$$

Then

$$QZ = PTT^{-1}V = PV = Y_r. \tag{17}$$

The elements of the T matrix will be determined by the researcher judgmentally, using the dictates of the theory or hypothesis that guided the research and data collection.

6. EMPIRICAL EVIDENCE

We will now present an empirical application of the technique to obtain estimates of individual parameters. The example comes from marketing. However, it is similar to many situations in other applied areas so that one can easily transfer the technique to other areas.

In marketing, one of the questions frequently raised is to find the consumer's response function either in a stable market or as due to some promotional activity of the company. The response function itself is measured in a variety of ways, but the most common practice is to take the frequency of purchases of a brand by the household during standard time intervals, such as a month or a quarter depending upon the purchase cycle. Generally, the purchase data are obtained from standard commercial panels, such as Market Research Corporation of America or The Chicago Tribune. The data used in this study came from a panel of about 400 families in the Chicago metropolitan area. The panel records purchases of a variety of grocery and personal care items on a continuous basis.

A sample of 154 families who bought a well-established convenience product for 60 or more consecutive times is used in the analysis. Each family's first 60 purchases of the product were grouped into 10 'trials' of six purchases each. The brand having the highest market share is the focal point of our analysis. The frequency of purchase of this brand at each 'trial' is calculated. This can vary from zero to six since each trial consists of six purchases.

The data matrix Y then consists of 10 rows, one for each trial session, and 154 columns, one for each family. The cell entry y_{ji} is the frequency of purchase of the brand by family i during trial j.

The first step in the analysis is to post-multiply Y with its transpose Y' which results in the cross-products matrix YY'. The post-multiplication vanishes the columns and so we have a 10×10 square symmetric cross-products matrix (Table 1). The diagonal values are the sums of squares and their sum represents the total variance in the system.

The square symmetric cross-products matrix is then resolved into its characteristic roots and vectors. The first three characteristic roots explain 90 per cent of the total variance, and, therefore, are considered adequate. Then taking the first three characteristic vectors and multiplying them with the square root of the characteristic roots, we obtain the A matrix, which is a 10×3 matrix containing three reference curves. In the usual terminology of factor analysis, we derive three factors which explain 90 per cent of total variance.

However, we have to rescale the values by pre-multiplying A with the scalar matrix $N^{-1/2}$ whose elements are $1/\sqrt{N}$. This is to standardize the sample size. The result is the P matrix which contains the rescaled reference curves. It is also a 10×3 matrix containing the same three reference curves as the matrix A. Table 2 gives the coefficients of each of the three reference curves.

The three reference curves are plotted on Figure 1 along with the mean curve. It will be noticed that at the aggregate level, the purchase frequency is constant with only slight deviations. More importantly, the first reference curve is identical in shape to the average curve, and it is also, therefore, a constant curve. The slight deviations in the parallelism of the two curves are at those points where the second or the third reference curves show substantial changes in their shapes. This supports the argument that the second and later reference curves act as 'correction terms' and provide a summary of the deviations from the average tendency. Finally, the excellent approximation to the

TABLE 1. CROSS-PRODUCTS MATRIX

		Trials									
		1	2	3	4	5	6	7	8	9	10
Trials	1	1430	1176	1157	1155	1126	1144	1131	1138	1176	1030
	2	1176	1306	1148	1149	1079	1116	1163	1082	1139	1031
	3	1157	1148	1318	1197	1122	1144	1150	1136	1182	1034
	4	1155	1149	1197	1387	1149	1216	1233	1208	1242	1095
	5	1126	1079	1122	1149	1327	1221	1192	1146	1220	1091
	6	1144	1116	1144	1216	1221	1516	1318	1317	1350	1200
	7	1131	1163	1150	1233	1192	1318	1495	1312	1328	1210
	8	1138	1082	1136	1208	1146	1317	1312	1505	1373	1216
	9	1176	1139	1182	1242	1220	1350	1328	1373	1595	1315
	10	1030	1031	1034	1095	1091	1200	1210	1216	1315	1377

TABLE 2. MEAN VALUES AND REFERENCE CURVE COEFFICIENTS FOR 10 TRIALS

Trial	Mean Value	Reference Curve 1	Reference Curve 2	Reference Curve 3
1	2.4025	2.702	.982	−.609
2	2.3060	2.639	.822	−.014
3	2.3116	2.687	.698	.013
4	2.3961	2.793	.370	.410
5	2.3311	2.709	.161	.298
6	2.5454	2.888	−.454	.530
7	2.4090	2.914	−.339	.460
8	2.3701	2.893	−.628	.026
9	2.5000	3.006	−.658	−.461
10	2.2532	2.695	−.766	−.871

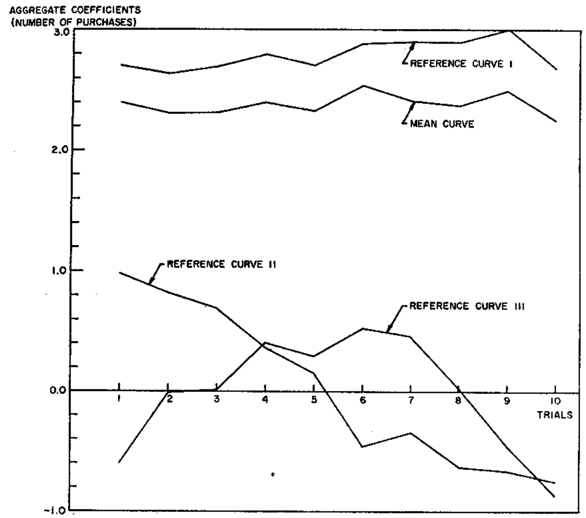

FIG. 1

average curve, more than 80 per cent of total variance explained by it alone, and the extremely high values compared to the other two curves suggest that a very large proportion of the sample consumers have the shape of their individual functions the same as the first reference curve. This also means that only a minority of consumers are likely to have functional relations which are similar to the second or the third reference curves.

The second reference curve possesses a negative slope and it seems to be a monotonically decreasing function of time (trials). Furthermore, it seems linear despite some fluctuations. It suggests that there are some consumers in the sample who constantly decrease or reciprocally increase the frequency of purchase of the brand as they continue to purchase the brand.

The third reference curve is a nonmonotonic function of time (trials). As such it seems to have a maximum near the center of the trials. It suggests that the sample contains some individuals who at first increased (or reciprocally decreased) the frequency of purchase of the brand, reached some maximum (minimum) level and then declined (increased) the frequency of purchase near the end of trials.

7. INDIVIDUAL PARAMETER ESTIMATES

More interesting are the coefficients s_{mi} for each individual in the sample. They determine his functional relation. There are 3 sets of coefficients for each

TABLE 3. PARAMETERS OF INDIVIDUAL FAMILIES: SOME EXAMPLES

Family No.	Reference Curves			
	I	II	III	
2	2.14	.28	−.71	
22	1.42	.72	.53	
25	1.86	−.15	−.06	
29	2.04	−.37	−.36	Type One Families
32	1.55	−.78	−.23	
154	1.68	.49	.41	
145	2.10	.03	−.38	
69	.66	2.31	.21	
138	1.02	−3.36	−1.43	
142	.53	1.27	−.03	Type Two Families
131	.96	−3.61	−.08	
130	1.00	4.07	.29	
68	1.27	−2.74	−.64	
95	.50	−.16	1.88	
55	.46	.55	2.49	
147	1.00	−.25	2.24	Type Three Families
103	.67	.46	−4.17	
53	1.40	−.86	−2.57	
1	.88	−2.11	3.42	
19	1.04	−.27	−1.21	
27	1.18	−.37	1.60	Mixed Type Families
50	1.76	−1.00	−1.50	
93	1.58	−1.60	−.36	
124	1.35	1.51	.47	

individual ($m=3$) since we have three reference curves. They are contained in the rescaled matrix V. Table 3 presents some illustrative families extracted from the V' matrix. Each row in that table reflects individual i's three coefficients.

The sample of families used in Table 3 can be broadly classified into four types of consumers: (i) Type one who have a functional relation similar to the first reference curve—constant frequency of purchases of a brand, (ii) Type two which is either monotonically increasing or decreasing function of trials, similar in form to the second reference curve, (iii) Type three consumers whose functional relation is nonmonotonic, and who either have a maximum or a minimum point near the center of the trial axis, and (iv) Type four consumers who have a combination of two or more reference curves which means that certain parts of their functional relation resemble one reference curve and other parts resemble some other reference curve.

We are thus able to estimate *and* summarize a large number of individual functions in a compact matrix. These estimates, furthermore, are least squares estimates and thus provide the best fit to the raw data. This can be seen by plotting individual consumer's observed frequencies over 10 trials and estimat-

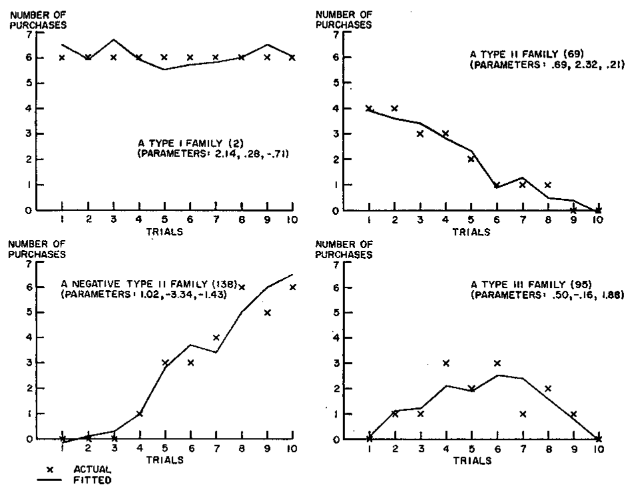

Fig. 2

ing his frequencies using the method described earlier. In Figure 2, we take four consumers (from the first three types) and plot their observed frequencies. Furthermore a curve is fitted in each of the plots based on the $\hat{y}_{ji}$ obtained from the a_{jm} and s_{mi} values. In all the cases, it will be seen that the fit is exceptionally good.

The first consumer (family 2) is a type one consumer and his functional relation is constant or independent of trials. His fitted curve is most similar to the first reference curve.

The second consumer (family 69) is a type two consumer and his functional relationship is monotonically decreasing over trials. His fitted curve is almost identical to the second reference curve.

The third consumer (family 138) is a type two consumer also but his functional relation is monotonically increasing over trials. His curve is also similar to the second reference curve except that the sign is opposite.

Despite the fact that the second and third consumers are reciprocally similar, there are several subtle differences which are also summarized in their s_{mi} coefficients. First, it will be seen from observed frequencies that family 138 is a slightly heavier buyer compared to family 69. This difference in level is incorporated in the s_{1i} coefficients: 1.02 and .66 respectively. Second, the former shows a rapid approach from zero frequency in trial 3 to the maximum frequency in trial 8: his slope is therefore greater than the latter. This is reflected in the larger coefficient for the s_{2i} score in the case of family 138. Finally, he shows greater tendency toward mononotonic relation at several points on the trial axis than the other consumer. This tendency is summarized in a greater score on s_{3i} for him.

Finally, the fourth consumer (family 95) shows a clear nonmonotonic relation. He at first increases the frequency of purchase of the brand, and after reaching the peak frequency near the central trials, he reduces the frequency to the zero level near the 10th trial. His fitted curve resembles most with the third reference curve. His s_{mi} coefficients also show him to be a clear type three consumer.

We have chosen only these four consumers to show the usefulness of the technique. The V' matrix in addition to the other illustrative families in Table 3 summarizes information for all the consumers in the sample.

8. IMPLICATIONS

There are several important implications which follow from the method discussed here. First, given an aggregate set of data on several individuals, it is no longer necessary to ignore the problem of averaging functional relations which are different in shape, and more importantly, in form. Using the factor analytic procedures, we not only disaggregate the aggregate function into dimensions or components but in the process classify homogeneous functions into one dimension or type. In the empirical example, we disaggregated consumers into at least three distinct types: those having constant frequency over time, those having monotonically increasing or decreasing ferquency, and those having nonmonotonic relation.

Second, each individual's parameter of the functional relation is separately estimated and summarily stored as his s_{mi} scores. Thus we not only estimate individual parameters but also compactly store them. The latter is important when we are dealing with large samples as we usually do in many applied disciplines of social sciences.

Third, the method can be used in estimating parameters for both theoretically-based or empirically-derived functions. Earlier, we gave examples of several exact functions (linear and nonlinear) which could be transformed into the linear postulate. Each of them could have been a function dictated by theory. More importantly, our empirical example was strictly data-based and there were no functions a priori set out based on some theory. With the information explosion we are witnessing today, methods which estimate parameters of strictly empirical data become extremely important. This is furthermore true in most of the social sciences which are still not mature sciences similar to physics, for example, and must therefore, rely more on empirical observations and work inductively.

Although the factor analytic method was described and used only in the realm of time-based functions, there is no need to restrict it. It could be extended to any independent variable x which has been discretely observed. For example, the author has used the method to disaggregate the average revenue (demand) function where the independent variable is discrete price points. Similarly, the method can be extended to dichotomous variables (one-zero) as well to multichotomous variables. Some preliminary work in this direction has also been completed with promising results [6].

There is, however, one problem in the factor analytic method described here. The parameters of individual functions (s_{mi}) are *derived* parameters, and they are, to that extent, a transformation of the more primitive parameters. This means that their interpretation and communication need a different approach than what is traditionally done in curve fitting and regression analyses. However, this is a problem with any new techniques and can only be resolved with passage of time and appropriate diffusion of the technique. Furthermore, the technique, in the process of generating derived parameters, does perform the two vital roles: (i) the parameters estimate and give the best fit to the observed data for each individual, and (ii) relative comparison of two or more individuals using the parameters is very possible since a distribution of the parameters can be easily obtained.

REFERENCES

[1] Eckart, C., and Young, G., "The approximation of one matrix by another of lower rank," *Psychometrika*, 1 (1936) 211–218.

[2] Estes, William K., "The problem of inference from curves based on group data," *Psychological Bulletin*, 53 (1956) 134–140.

[3] Merrill, M., "The relationship of individual growth to average growth," *Human Biology*, 3 (1931) 37–70.

[4] Rao, C. R., "Some statistical methods for comparison of growth curves," *Biometrics*, 14 (1958) 1–17.

[5] Ross, John, "Mean performance and the factor analysis of learning data," *Psychometrika*, 29 (1964) 67–73.

[6] Sheth, Jagdish N. *A Behavioral and Quantitative Investigation of Brand Loyalty* (unpublished Ph.D. Dissertation, University of Pittsburgh, 1966).

[7] Tucker, Ledyard R., "Determination of parameters of a functional relation by factor analysis," *Psychometrika*, 23 (1958) 19–23.

[8] Tucker, Ledyard R., "Learning theory and multivariate experiment: illustration by determination of generalized learning curves," in R. Cattell (ed.), *Handbook of Multivariate Experimental Psychology* (Chicago: Rand McNally, 1966), pp. 476–501.

[9] Young, G. and Householder, A. S., "Matrix approximation and latent roots," *American Mathematical Monthly*, 45 (1938) 165–171.