QUADRATIC FACTOR ANALYSIS: LINEAR DECODING OF THE HIGHER DATA MOMENTS

William W. Rozeboom and J. Jack McArdle
University of Alberta — University of Virginia

ABSTRACT

Quadratic factor analysis extends the logic of classical linear factoring for latent sources to analysis of all data moments through the 4th-order. Given the model presumptions, this yields identification of all factor moments through the 4th-order and from there discloses, inter alia, whether any of the data variables' orthothodoxly recovered common factors are in fact quadratic functions of the others, or nearly so.

Key words: Quadratic factoring. Higher-moment analysis. Nonlinear factor analysis.

This research has been supported by the National Sciences and Engineering Research Council of Canada through grant No. A1054 to the senior author.

Requests for reprints may be sent either to Wm. W. Rozeboom, Dept. of Psychology, Univ. of Alberta, Alberta, Canada T6G 2E9, or to J. Jack McArdle, Dept. of Psychology, Gilmore Hall, Univ. of Virginia, Charlottesville, VA, USA 22901.

June, 1987

QUADRATIC FACTOR ANALYSIS: LINEAR DECODING OF THE HIGHER DATA MOMENTS

William W. Rozeboom
University of Alberta

and

J. Jack McArdle
University of Virginia

Although linear factor analysis traditionally operates upon only the 2nd-order central moments (i.e. covariances) of multivariate data arrays, it has long been known that higher data moments also contain potentially useful information about the data's common sources. Yet apart from Latent Structure Analysis (see Lazersfeld, 1959; Lazersfeld & Henry, 1968), which has been developed primarily for treatment of binary variables and is severely limited in the complexity it can assimilate, few efforts have yet been made to interpret data moments higher than covariances--possibly because one might expect their analysis to require a mathematics far less tractable than the linear algebra which has proved so effective for analysis of covariance structures.

It turns out, however, that just as linear algebra can nicely handle curvilinear functions whose parameterizations are linear, so can the algorithms developed by linear factor analysis and more recently linear causal modelling informatively decompose data moments of all orders. (See Kenny & Judd, 1984, for solution of a restricted special case; Mooijaart, 1985, on positioning of factor axes by appeal to 3rd-order moments; and Bentler, 1983, p. 496f., for an overview of the generic moment model which does not, however, develop any solution practicalities.) We shall here set out the theory and computational praxis for inclusion of 3rd- and 4th-order data moments in the analysis. (Extension to even higher moments is clearly premature at this time.) It seems natural to call this procedure <u>Quadratic Factor Analysis</u>, or "quad-factoring" for short.

In brief, quad-factoring of data on an array $\underline{Y} = \{\underline{y}_i\}$ of metrical scales supplements the variables in $\underline{Y}$ by their pairwise products $\{y_{ij} = \underline{y}_i\underline{y}_j\}$, and observes that any orthodox linear common-factor model for the 1st-level array $\underline{Y}$ entails a corresponding linear model for the expanded (2nd-level) array as well. Just as traditional factoring extracts model parameters from the 1st-level data covariances, so does quad-factoring solve the quad-moments counterpart of covariances--namely, the 1st-level variables' moments through the 4th order--for parameters in the factor model's quadratic extension. In principle, quadratic factoring should disclose the same common-factor loadings and uniquenesses for the data variables as does traditional 1st-level factoring. But when the quad-factoring model premises are not violated too outrageously, it should identify communalities and weak common factors with greater precision than does 1st-level analysis. In particular, it resolves uniqueness ambiguities in 1st-level factoring such as arise from doublet factors. Even more importantly, quad-factoring recovers not merely common-factor covariances but all factor moments through the 4th order. Theories of what we can gain from this higher-moment information still remain largely underdeveloped. But one major prospect is detection of nonlinearities in the functions by which our data variables arise from their real underlying sources (see p. 12, below). And it can strongly ajudicate conjectures (e.g., Gangestad & Snyder, 1985) that the factors diagnosed by certain test items are dichotomous.

Terminology and model presumptions.

The presumptions of quadratic factoring are stronger than traditional in factor analysis, but only modestly so. We begin with any standard metrical data array, that is, the joint distribution in some sample population $\underline{P}$ on an $\underline{n}$-tuple $\underline{Y} = \langle \underline{y}_1,\ldots,\underline{y}_n \rangle$ of metrical output variables. (When relevant, read $\underline{Y}$ and other tuples of variables as column vectors of their components.) We shall not here address sampling issues, so for simplicity we equate the arithmetic mean, $\underline{m}_x$, of any measure $\underline{x}$ distributed in $\underline{P}$ with $\underline{x}$'s expectation $\mathcal{E}[\underline{x}]$ in the population sampled. It is convenient to scale all the $\underline{Y}$-variables--call these our 1st-level data variables--to have zero means in $\underline{P}$; but variance normalization is optional, and eventually we allow 1st-level centering to be waived as well. Next, define (proper) 2nd-level data variables $\underline{Y}^* = \{y_{ij}: i = 1,\ldots,n;\ j = i,\ldots,n\}$ to be the $n(n+1)/2$ pairwise product variables $y_{ij} =_{def} y_i y_j$ $(i \leq j)$ such that each subject's score on y_{ij} is the product of his scores on y_i and y_j. Each 1st-level variable y_i, too, can be viewed as a special 2nd-level variable $y_i = y_{0i} = y_0 y_i$ where y_0 is the unit variable on which, by definition, all scores are unity. (We shall designate the unit variable by a variety of letters, but always with a subscript of 0.) When $\underline{Y}$-scores are known for members of $\underline{P}$, the same is evidently true for all product-variables in $\underline{Y}^*$. It will be important to leave each $\underline{Y}^*$-variable y_{ij} in the metric defined for it by its constituents y_i and y_j. That is, neither the mean nor variance of y_{ij} is adjusted beyond what is imposed by choice of scales for y_i and y_j.

Since we shall have repeated need, with variations, for the notation just introduced, we had best take pains to set this out in full generality. Let $\underline{X} = \langle x_s, x_{s+1}, \ldots, x_n \rangle$ be any $(n-s+1)$-tuple of variables indexed consecutively from a starting index s. (We shall use only $s = 0$ and $s = 1$.) Then the (full) quadratic development $\underline{X}^{@}$ of X is the $(n-s+1)^2$-tuple of pairwise product-variables

$$\underline{X}^{@} = \{x_{ij}: x_{ij} = x_i x_j;\ i,j = s,\ldots,n\} \quad ;$$

while the (bare) quadratic development $\underline{X}^*$ of $\underline{X}$ is the $(n-s+1)(n-s+2)/2$-tuple that remains of $\underline{X}^{@}$ when all x_{ij} in which $i > j$ are deleted from it, namely,

$$\underline{X}^* = \{x_{ij}: x_{ij} = x_i x_j;\ i = s,\ldots,n;\ j = i,\ldots,n\} \quad .$$

(Since $x_{ij} = x_i x_j = x_j x_i = x_{ji}$, $\underline{X}^{@}$ contains $(n-s+1)(n-s)/2$ duplications which are eliminated in $\underline{X}^*$. Our practical work will be with $\underline{X}^*$; but $\underline{X}^{@}$ yields the tidier algebraic theory.) Secondly, for any tuple $\underline{X} = \langle x_1,\ldots,x_n \rangle$ of variables with starting index 1, we write $\underline{X}_0$ for $\underline{X}$ preceded by the unit variable x_0. That is,

$$\underline{X}_0 =_{def} \langle x_0, \underline{X} \rangle = \langle x_0, x_1, \ldots, x_n \rangle$$

wherein all scores on x_0 are unity. Then the full/bare quadratic development of $\underline{X}_0$

includes not only the full/bare quadratic development of $\underline{X}$ but 1st-level variables $\underline{X}$ as well. Specifically,

$$\underline{X}_0^* = \langle \underline{x}_{00}, \underline{x}_{01}, \ldots, \underline{x}_{0n}, \underline{X}^* \rangle = \langle \underline{x}_0, \underline{X}, \underline{X}^* \rangle$$

where $\underline{x}_{00} = \underline{x}_0\underline{x}_0 = \underline{x}_0$. And $\underline{X}_0^{@}$ similarly includes $\underline{x}_0$ and $\underline{X}$ along with $\underline{X}^{@}$. So given a tuple of variables with starting index 1, we can refer to just their proper 2nd-level products as $\underline{X}^*$ or (with duplications) as $\underline{X}^{@}$, and to their 1st-and-2nd-level ensemble combined along with $\underline{x}_0$ as $\underline{X}_0^*$ or $\underline{X}_0^{@}$.

The matrix $\underset{\sim}{C}_{YY}$ of covariances among 1st-level output variables $\underline{Y}$ on which linear data analysis traditionally operates comprises the 2nd-order central moments of the $\underline{Y}$-distribution in $\underline{P}$. That is, depending on whether we distinguish $\underline{P}$ from the population sampled by $\underline{P}$, $[\underset{\sim}{C}_{YY}]_{ij}$ either equals $\mathcal{E}[(\underline{y}_i - \mathcal{E}[\underline{y}_i])(\underline{y}_j - \mathcal{E}[\underline{y}_j])]$ or is a sampling estimate thereof. Quad-factoring, however, works with 2nd-order moments (of the product-variables) that are not generally centered. So for any two tuples of variables $\underline{X} = \langle \ldots, \underline{x}_\alpha, \ldots \rangle$ and $\underline{Z} = \langle \ldots, \underline{z}_\beta, \ldots \rangle$ (not necessarily $\underline{X} \neq \underline{Z}$), we shall write $\underset{\sim}{M}_{XZ}$ or $\underset{\sim}{M}(\underline{X},\underline{Z})$ for the matrix whose $\alpha\beta$ th element $[\underset{\sim}{M}_{XZ}]_{\alpha\beta}$ is the mean product of $\underline{x}_\alpha$ and $\underline{z}_\beta$ in whatever population $\underline{P}$ is at issue. That is, under our simplifying identification of sample means with population expectations, $[\underset{\sim}{M}_{XZ}]_{\alpha\beta} = \mathcal{E}[\underline{x}_\alpha \underline{z}_\beta]$. This notation does **not** presume that the explicit index α of $\underline{x}_\alpha$ in $\underline{X}$ or β of $\underline{z}_\beta$ in $\underline{Z}$ is necessarily that variable's count-position in its tuple—cf. cases $\underline{X}_0 = \langle \underline{x}_0, \underline{x}_1, \ldots \rangle$ and $\underline{X}^* = \langle \ldots, \underline{x}_{ij}, \ldots \rangle$. Rather, $[\underset{\sim}{M}_{XZ}]_{\alpha\beta}$ is the element of $\underset{\sim}{M}_{XZ}$ in the row headed by $\underline{x}_\alpha$ and column headed by $\underline{z}_\beta$. In particular, for any doubly indexed array $\underline{X}^*$, $[\underset{\sim}{M}_{X^*X^*}]_{hi,jk} = \mathcal{E}[\underline{x}_{hi}, \underline{x}_{jk}] = \mathcal{E}[\underline{x}_h\underline{x}_i\underline{x}_j\underline{x}_k]$.

Because our notation for tuples of 2nd-level variables produces visual monstrosities and typesetters' nightmares when used as subscripts in traditional formulas for moment arrays, we shall henceforth treat $\underset{\sim}{m}$ (denoting a vector of means), $\underset{\sim}{C}$ (denoting a covariance matrix), and $\underset{\sim}{M}$ (denoting a matrix of uncentered 2nd-order moments) notationally as functions of the variables whose moments are at issue.

Thus $\underset{\sim}{m}_X$ and $\underset{\sim}{M}_{XZ}$ will generally be written as $\underset{\sim}{m}(\underline{X})$ and $\underset{\sim}{M}(\underline{X},\underline{Z})$, respectively.

Whenever $\underset{\sim}{M}$ is a matrix whose rows and columns are doubly indexed, we shall say that $\underset{\sim}{M}$ is <u>quad-symmetric</u> iff $[\underset{\sim}{M}]_{hi,jk} = [\underset{\sim}{M}]_{h'i',j'k'}$ whenever these terms are both well-defined elements of $\underset{\sim}{M}$ in which $\langle \underline{h}',\underline{i}',\underline{j}',\underline{k}'\rangle$ is a permutation of $\langle \underline{h},\underline{i},\underline{j},\underline{k}\rangle$. Clearly, $\underset{\sim}{M}(\underline{X}_0^{\circledast},\underline{X}_0^{\circledast})$ and $\underset{\sim}{M}(\underline{X}_0^*,\underline{X}_0^*)$ are quad-symmetric.

For any array of 1st-level data variables $\underline{Y} = \langle \underline{Y}_1,\ldots,\underline{Y}_n\rangle$, if $\underline{Y}^*$ is the bare quadratic development of $\underline{Y}$ as defined above, and $\underline{Y}_0^*$ is the bare quadratic development of $\underline{Y}$'s extension $\underline{Y}_0 = \langle \underline{Y}_0,\underline{Y}\rangle$ to include the unit variable, the 2nd-order moment matrix $\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*)$ of $\underline{Y}_0^*$ partitions as

$$\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*) = \begin{bmatrix} 1 & & \underline{\text{sym}} \\ \underset{\sim}{m}(\underline{Y}) & \underset{\sim}{M}(\underline{Y},\underline{Y}) & \\ \underset{\sim}{m}(\underline{Y}^*) & \underset{\sim}{M}(\underline{Y}^*,\underline{Y}) & \underset{\sim}{M}(\underline{Y}^*,\underline{Y}^*) \end{bmatrix} \qquad (\, \underline{Y}_0^* = \langle \underline{Y}_0,\underline{Y},\underline{Y}^*\rangle \,) \quad ,$$

wherein "<u>sym</u>" signifies symmetry. This makes clear that all moments of $\underline{Y}$ through the 4th order are contained in $\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*)$. The 1st-order moments are in vector $\underset{\sim}{m}(\underline{Y})$ (= $\underset{\sim}{0}$ under centered scaling of $\underline{Y}$); the 2nd-order moments are in $\underset{\sim}{M}(\underline{Y},\underline{Y})$ (= $\underset{\sim}{C}_{YY}$ under centered scaling) and also, rearranged as a vector, in $\underset{\sim}{m}(\underline{Y}^*)$; the 3rd-order moments are in $\underset{\sim}{M}(\underline{Y}^*,\underline{Y})$; and the 4th-order moments are in $\underset{\sim}{M}(\underline{Y}^*,\underline{Y}^*)$.

The point now to be developed is that when all $\underline{Y}$-moments through the 4th order--call these the "quad-moments" of $\underline{Y}$--are so treated as the 2nd-order moment matrix of $\underline{Y}_0$'s quadratic development, we can analyze $\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*)$ for information about $\underline{Y}^*$'s factor composition by the very same linear models that have traditionally worked so well on 1st-level covariances. We retain the classic premise that each 1st-level data variable is the sum of a common part and unique residual which we find convenient to construe as a psychometric "true-part" and "error," respectively. Specifically, we posit

$$\underline{Y}_i = \underline{t}_i + \underline{e}_i \qquad (\, \underline{i} = 1,\ldots,\underline{n}) \quad , \tag{1}$$

with "error" characterized by one essential distributional constraint and two auxillary ones that are expository conveniences easily waived in computational practice:

The basic quad-factoring error premise. First-level residuals $\underline{E} = \langle \underline{e}_1, \ldots, \underline{e}_n \rangle$ in (1) have zero expectations, and are distributed independently of one another and of all true-parts $\underline{T} = \langle \underline{t}_1, \ldots, \underline{t}_n \rangle$. (See Appendix A, Note, for

Strong error-model addenda [optional]. The marginal distribution of each $\underline{e}_i$ in (1) has the same skew and kurtosis as a Normal distribution.

Meanwhile, 1st-level true-parts $\underline{T}$ are presumed to be linear combinations of a smaller number of common factors which in turn may or may not be different linear/nonlinear functions of a still-smaller number of substantively distinct common sources. This 1st-level factor model entails a well-behaved factor model for the 2nd-level data variables as well, or rather for their true-parts. The 2nd-level error model for quad-factoring, however, is more complicated than its 1st-level counterpart; and its theory is our lead-off concern.

Second-level error theory.

Given psychometric model (1) for 1st-level variables $\underline{Y}$, each 2nd-level variable $\underline{y}_{ij} = \underline{y}_i \underline{y}_j = (\underline{t}_i + \underline{e}_i)(\underline{t}_j + \underline{e}_j)$ in $\underline{Y}^*$ has true-part/error composition

$$\underline{y}_{ij} = \underline{t}_{ij} + \underline{e}_{ij} \qquad (2.1)$$

where

$$\underline{t}_{ij} =_{\text{def}} \underline{t}_i \underline{t}_j \,, \qquad \underline{e}_{ij} =_{\text{def}} \underline{t}_i \underline{e}_j + \underline{e}_i \underline{t}_j + \underline{e}_i \underline{e}_j \,. \qquad (2.2)$$

For $\underline{j} = 0$ we stipulate

$$\underline{t}_0 = 1 \,, \qquad \underline{e}_0 = 0 \,,$$

to yield

$$\underline{t}_{0j} = \underline{t}_j \,, \quad \underline{e}_{0j} = \underline{e}_j \,, \qquad (\, \underline{j} = 0,1,\ldots,\underline{n} \,)$$

and hence $\underline{y}_{0j} = \underline{t}_{0j} + \underline{e}_{0j}$ for each $\underline{j} = 0,1,\ldots,\underline{n}$. So if we write

$$\underline{E}_0^+ =_{def} \{\underline{e}_{ij}: \underline{i} = 0,1,\ldots,\underline{n};\ \underline{j} = \underline{i},\ldots,\underline{n}\} \quad ,$$

(1) becomes a fragment of

$$\underline{Y}_0^* = \underline{T}_0^* + \underline{E}_0^+ \tag{3}$$

wherein $\underline{Y}_0^*$ and $\underline{T}_0^*$ are the bare quadratic developments of 1st-level data variables $\underline{Y}_0 = \langle \underline{y}_0, \underline{Y}\rangle$ and their true-parts $\underline{T}_0 = \langle \underline{t}_0, \underline{T}\rangle$, and $\underline{E}_0^+$ comprises the corresponding residuals. (Note from (2.2), however, that $\underline{E}_0^+$ is not quadratic development $\underline{E}_0^*$ of $\underline{E}_0$. Rather, $\underline{E}_0^*$ is just one of three components in $\underline{E}_0^+$. And $\underline{e}_0$ is constant at zero rather than at unity. So the $\underline{e}$-variables are exceptions to the subscript conventions we have adopted for non-error variables.) $\underline{T}_0^*$ and $\underline{E}_0^+$ are respectively the true-part and error components of 2nd-level data variables $\underline{Y}_0^*$; and (3)'s additivity insures that data quad-moments $M(\underline{Y}_0^*, \underline{Y}_0^*)$ likewise decompose as a sum of true-part and error terms.

From (2.1), it is evident that each 2nd-order moment $[\mathbf{M}(\underline{Y}^*, \underline{Y}^*)]_{hi,jk}$ of $\underline{Y}_0$'s quadratic development $\underline{Y}_0^*$ has composition $\mathcal{E}[\underline{y}_{hi}\underline{y}_{jk}] = \mathcal{E}[(\underline{t}_{hi} + \underline{e}_{hi})(\underline{t}_{jk} + \underline{e}_{jk})] = \mathcal{E}[\underline{t}_{hi}\underline{t}_{jk}] + \mathcal{E}[\underline{e}_{hi}\underline{t}_{jk}] + \mathcal{E}[\underline{e}_{hi}\underline{t}_{jk}] + \mathcal{E}[\underline{e}_{hi}\underline{e}_{jk}]$. That is,

$$\mathbf{M}(\underline{Y}_0^*, \underline{Y}_0^*) = \mathbf{M}(\underline{T}_0^*, \underline{T}_0^*) + \mathbf{M}(\underline{T}_0^*, \underline{E}_0^+) + \mathbf{M}'(\underline{T}_0^*, \underline{E}_0^+) + \mathbf{M}(\underline{E}_0^+, \underline{E}_0^+) \quad . \tag{4}$$

Unlike error covariances in 1st-level data, $\mathbf{M}(\underline{E}_0^+, \underline{E}_0^+)$ is not altogether diagonal nor is $\mathbf{M}(\underline{T}_0^*, \underline{E}_0^+)$ wholly zero. Even so, under the quad-factoring error premises these are identifiable from the 1st-level uniquenesses (error variances) and observed 1st-level covariances. For parameters, let us write

$$\underline{u}_i =_{def} \mathcal{E}[\underline{e}_i^2] \ , \qquad \underline{c}_{ij} =_{def} \begin{cases} \mathcal{E}[\underline{y}_i^2] = \mathcal{E}[\underline{t}_i^2] + \underline{u}_i & \text{if } \underline{i} = \underline{j} \\ \mathcal{E}[\underline{y}_i\underline{y}_j] = \mathcal{E}[\underline{t}_i\underline{t}_j] & \text{if } \underline{i} \neq \underline{j} \end{cases} \quad ,$$

noting that $\underline{u}_0 = 0$, $\underline{c}_{00} = 1$, and $\underline{c}_{0j} = \mathcal{E}[\underline{y}_j]$ for index 0. That is, for $\underline{i},\underline{j} = 0,\ldots,\underline{n}$,

$$\underline{c}_{ij} = [\mathbf{M}(\underline{Y}_0, \underline{Y}_0)]_{ij} \ , \qquad \underline{u}_i = [\mathbf{M}(\underline{Y}_0, \underline{Y}_0) - \mathbf{M}(\underline{T}_0, \underline{T}_0)]_{ii} \ .$$

For centered $\underset{\sim}{Y}$, c_{ij} equals data covariance $[C_{YY}]_{ij}$ for $i,j > 0$; and u_i is the traditional "uniqueness" of data variable y_i.

In the strong error model, the 1st-level $\{u_i\}$ are the only unknown error parameters. But to waive the strong error model's Normality assumption, we also require parameters for the raw (unstandardized) error skew and kurtosis. So for these we shall write

$$u_i^{[3]} =_{def} \mathcal{E}[e_i^3] , \qquad u_i^{[4]} =_{def} \mathcal{E}[e_i^4] ,$$

for $i = 1,\ldots,n$. In the strong error model, $u_i^{[3]} = 0$ and $u_i^{[4]} = 3u_i$.

Finally, since separation of the three error matrices in (4) serves no purpose, we put

$$Q(E_0^+) =_{def} M(T_0^*,E_0^+) + M'(T_0^*,E_0^+) + M(E_0^+,E_0^+) ,$$

whence (4) simplifies to

$$M(Y_0^*,Y_0^*) = M(T_0^*,T_0^*) + Q(E_0^+) . \qquad (4')$$

Because (4') is the error/true-part decomposition of Y's <u>bare</u> quadratic development Y_0^*, the elements $[Q(E_0^+)]_{hi,jk}$ of $Q(E_0^+)$ are under index constraint $h \leq i$ and $j \leq k$. To avoid this expository nusiance, we shall speak instead of $Q(E_0^+)$'s full-quadratic-development counterpart

$$Q(E_0^{\oplus}) =_{def} M(Y_0^{\oplus},Y_0^{\oplus}) - M(T_0^{\oplus},T_0^{\oplus}) \qquad (5)$$

and write $q_{hi,jk}$ for an arbitrary element thereof. That is,

$$q_{hi,jk} =_{def} [Q(E_0^{\oplus})]_{hi,jk}$$

for all $h,i,j,k = 0,1,\ldots,n$, with $q_{hi,jk}$ being also the $\langle hi,jk \rangle$th element of $Q(E_0^+)$ iff $0 \leq h \leq i \leq n$ and $0 \leq j \leq k \leq n$.

In Appendix A, we show that each element of $Q(E_0^{\oplus})$ is identical up to permutation of its four 1st-order indices with some subscript instantiation in

$$
\begin{aligned}
\underset{\sim}{g}_{0i,0j} &= 0 \,, && (6)\\
\underset{\sim}{g}_{hi,jk} &= 0 \qquad (\underline{h},\underline{i},\underline{j},\underline{k} \text{ all distinct}) \,,\\
\underset{\sim}{g}_{ii,jk} &= \underline{c}_{jk}\underline{u}_i \qquad (\underline{i},\underline{j},\underline{k} \text{ all distinct}) \,,\\
\underset{\sim}{g}_{0i,ii} &= \underline{u}_i^{[3]} \qquad (\text{centered } \underline{y}_i) \,,\\
&= 0 \text{ in the strong error model} \,,\\
\underset{\sim}{g}_{hi,ii} &= 3\underline{c}_{hi}\underline{u}_i \qquad (0 < \underline{h} \neq \underline{i};\ \text{centered } \underline{y}_i) \,,\\
\underset{\sim}{g}_{ii,jj} &= (\underline{c}_{ii} - \underline{u}_i)\underline{u}_j + (\underline{c}_{jj} - \underline{u}_j)\underline{u}_i + \underline{u}_i\underline{u}_j \\
&= \underline{c}_{ii}\underline{u}_j + \underline{c}_{jj}\underline{u}_i - \underline{u}_i\underline{u}_j \qquad (\underline{i} \neq \underline{j}) \,,\\
\underset{\sim}{g}_{ii,ii} &= 6(\underline{c}_{ii} - \underline{u}_i)\underline{u}_i + \underline{u}_i^{[4]} \\
&= 6\underline{c}_{ii}\underline{u}_i - 3\underline{u}_i^2 \text{ in the strong error model} \,.
\end{aligned}
$$

The elements of $\underset{\sim}{G}(\underline{E}_0^{\oplus})$ are indifferent to all permutations of their 1st-order indices, which is to say not merely that $g_{hi,jk} = g_{jk,hi}$ and $g_{hi,jk} = g_{ih,jk} = g_{hi,kj}$, but also $g_{hi,jk} = g_{hj,ik} = g_{ji,hk}$. When all $\underline{Y}$-variables have standard scales, i.e. zero means and unit variances, $\underline{c}_{ii} = \underline{c}_{jj} = 1$ in the formulas for $g_{ii,jj}$ and $g_{ii,ii}$.

Given the 1st-level data covariances, it is straightforward to produce 2nd-level error matrix $\underset{\sim}{G}(\underline{E}_0^{+})$ from (6) either algebraically as a function of $\underset{\sim}{C}_{YY}$ and the uniqueness parameters or as a numerical estimate derived from $\underset{\sim}{C}_{YY}$ and a provisional solution for the latter. And the solution algorithm can iterate estimation of $\underset{\sim}{G}(\underline{E}_0^{+})$ just as 1st-level factor analysis has traditionally iterated uniqueness estimation when high-grade results are wanted. Whatever our provisional solution for $\underset{\sim}{G}(\underline{E}_0^{+})$, this gives a corresponding estimate of the 2nd-level true-parts' moment matrix $\underset{\sim}{M}(\underline{T}_0^{*},\underline{T}_0^{*}) = \underset{\sim}{M}(\underline{Y}_0^{*},\underline{Y}_0^{*}) - \underset{\sim}{G}(\underline{E}_0^{+})$ which embeds the 1st-level true-parts' covariances and can be searched for interpretable structure by standard methods of matrix decomposition. But we have not yet considered what is there to be found.

<u>Second-level factor patterns</u>.

As already declared, we posit traditional factor model

$$\underline{t}_i = \sum_{j}^{r} a_{ij}\underline{f}_j \qquad (\underline{i} = 1,\ldots,\underline{n};\ \underline{r} < \underline{n}) \tag{7}$$

for our centered 1st-level data variables' true-parts, with the number $\underline{r}$ of 1st-level factors $\underline{f}_1,\ldots,\underline{f}_r$ an open parameter. It is notorious that this decomposition of $\underline{T}$ is flagrantly nonunique, not merely under factor rotation but even in its dimensionality albeit we orthodoxly choose $\underline{r}$ as small as is compatable with good reproduction of the data covariances. Even so, for any specific choice of factors $\underline{F} = \langle \underline{f}_1,\ldots,\underline{f}_r \rangle$ in (7) there is a corresponding factor decomposition of the true-parts $\underline{T}^* = \langle \ldots,\underline{t}_{ij},\ldots \rangle$ of 2nd-level data variables $\underline{Y}^*$. For it is an obvious consequence of (2.2) and (7) that

$$\underline{t}_{hi} = \left(\sum_{j=1}^{r} \underline{a}_{hj}\underline{f}_j\right)\left(\sum_{j=1}^{r} \underline{a}_{ik}\underline{f}_k\right) = \sum_{k=1}^{r}\sum_{j=1}^{r} \underline{a}_{hj}\underline{a}_{ik}\underline{f}_j\underline{f}_k \quad . \tag{8}$$

Let $\underline{F}^* = \langle \ldots,\underline{f}_{jk},\ldots \rangle$ be the bare quadratic development of 1st-level factors $\underline{F}$, i.e.

$$\underline{f}_{jk} =_{def} \underline{f}_j\underline{f}_k \quad (\underline{j} = 1,\ldots,\underline{r};\ \underline{k} = \underline{j},\ldots,\underline{r}) \quad .$$

Theory will soon prefer that we extend (8) into a 2nd-level factor pattern for the combined $\underline{T}^*_0 = \langle \underline{t}_0,\underline{T},\underline{T}^* \rangle$ upon $\underline{F}^*_0 = \langle \underline{f}_0,\underline{F},\underline{F}^* \rangle$. But for openers let us consider just the pattern of $\underline{T}^*$ upon the proper 2nd-level factors $\underline{F}^*$. Noting that $\underline{f}_{jk}$ occurs twice in (8) if $\underline{j} \neq \underline{k}$, once as $\underline{f}_j\underline{f}_k$ and again as $\underline{f}_k\underline{f}_j$, we can rewrite (8) as

$$\underline{t}_{hi} = \sum_{j=1}^{r}\sum_{k=j}^{r} \underline{a}_{hi,jk}\underline{f}_{jk} \quad (\underline{h} = 1,\ldots,\underline{n};\ \underline{i} = \underline{h},\ldots,\underline{n}) \tag{9.1}$$

wherein

$$\underline{a}_{hi,jk} =_{def} \begin{cases} \underline{a}_{hj}\underline{a}_{ik} + \underline{a}_{hk}\underline{a}_{ij} & \text{if } \underline{j} \neq \underline{k} \\ \underline{a}_{hj}\underline{a}_{ik} & \text{if } \underline{j} = \underline{k} \quad . \end{cases} \tag{9.2}$$

Most noteworthy about (9) is simply its exhibiting how 2nd-level true-part variables $\underline{T}^*$ inherit a linear factor composition from any that holds for their 1st-level generators $\underline{T}$. So this 2nd-level pattern, along with factor moments $\underset{\sim}{M}(\underline{F}^*,\underline{F}^*)$, should be recoverable from $\underset{\sim}{M}(\underline{T}^*,\underline{T}^*)$ by methods already familiar in 1st-level factoring. Indeed, the factor pattern in (9) appears even more strongly structured than is the 1st-level pattern from which it derives: Whereas the number-of-factors/number-of-data-variables ratio at the 1st level is $\underline{r}/\underline{n}$, at the 2nd level this is only

$\underline{r}(\underline{r}+1)/\underline{n}(\underline{n}+1) \simeq (\underline{r}/\underline{n})^2$. And 2nd-level variable $\underline{y}_{hi}$ has appreciable loading on 2nd-level factor $\underline{f}_{jk}$ only if $\underline{y}_h$ has appreciable loading on one of $\underline{f}_j$ or $\underline{f}_k$ while $\underline{y}_i$ loads appreciably on the other. So one might also anticipate that quad-factoring should identify simple-structure hyperplanes more sharply than 1st-level factoring usually achieves. Unhappily, our inquiry into this prospect suggests it to be largely illusory (cf. p. 22, below). But it still remains one incentive to explore quad-factoring's potential with some care.

Why bother?

Before grubbing into solution details, some motivation stronger than hopes for pretty hyperplanes seems called for: It is all very well to observe that 1st-level factor patterns entail 2nd-level ones. But if the latter are redundant with the former, what point might there be in seeking solutions at both levels? Our wisdom in this regard is still too nascent for a confident answer. But we foresee two ways in which this may well prove profitable.

One important prospect lies in the 1st-level/2nd-level pattern redundancy itself. It is well known that common-factoring seldom picks out one particular solution as pronouncedly superior to all alternatives. Solving for 1st- and 2nd-level patterns simultaneously under constraint (9.2) in principle yields results more strongly overdetermined, and hence more finely discriminating of what seems optimal, than 1st-level analysis alone can provide. In particular, enhanced overdetermination should enable quad-factoring to capture factors too weak for detection at just the 1st level. (How well this will work out in the teeth of sampling error and other real-data model violations remains to be seen; but the artificial-data studies summarized in Appendix D are mildly encouraging.)

Even more provocative is what quad-factoring can tell us about the 3rd- and 4th-order moments of the 1st-level factors. Identifying these higher factor moments is straightforward in principle: When our factoring of the 2nd-level variables rotates their true-parts' factor axes to positions and scalings on which the 2ndlevel pattern is related to the 1st-level one as (9) is to (7), then each 2nd-level factor is tagged as the product of two particular 1st-level factors (or as the square of one). And the mean product of 2nd-level factors $\underline{f}_{hi} = \underline{f}_h\underline{f}_i$ and $\underline{f}_{jk} = \underline{f}_j\underline{f}_k$, which we compute along with the factor pattern, is then a 4th-order moment $\mathcal{E}[\underline{f}_h\underline{f}_i\underline{f}_j\underline{f}_k]$ of the 1st-level factor distribution. More completely, analysis of combined-levels data variables $\underline{Y}_0^*$ gives us the array $\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)$ of all $\underline{F}$-moments through the 4th order. And that in turn diagnoses, inter alia, whether some of 1st-level factors $\underline{F}$ are themselves quadratic functions of the others, or nearly so.

There is nothing in the linearity of an orthodox 1st-level factor decomposition to preclude its being an artifact of what in reality is a curvilinear production of these outputs by their common causes. Specifically, (7) may well be a linear parameterization of some nonlinear determination

$$\underline{t}_i = \sum_{j=1}^{r} a_{ij}\phi_j(g_1,\ldots,g_s) \qquad (\underline{i} = 1,\ldots,\underline{n})$$

of the data variables' true-parts by certain sources $\underline{G} = \langle g_1,\ldots,g_s\rangle$ of which the more manifest factors $\underline{F} = \langle \underline{f}_1,\ldots,\underline{f}_r\rangle$ are various nonlinear composites $\{\underline{f}_j = \phi_j(\underline{G})\}$. (Cf. McDonald, 1962; Rozeboom, 1965, p. 523ff.) If so, Taylor-series expansion allows us to hope that many—with luck, most or all—of these $\phi_j(\underline{G})$ are approximated by quadratic functions of $\underline{G}$ closely enough to leave negligible residuals. (For example, $\underline{f}_1,\ldots,\underline{f}_5$ might be centerings of quadratic functions g_1, g_2, g_1^2, g_2^2, g_1g_2, respectively, of just two real sources $G = \langle g_1,g_2\rangle$.) If so, whatever 1st-level factors of $\underline{Y}$ are quadratic functions of the others will lie in the quadratic space of $\underline{Y}_0$'s true-part $\underline{T}_0$, and can be identified as such from $\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)$.

__Fragments of the theory of quadratic spaces__ (__précis__).

As you might expect, certain technicalities in the mathematics of quadratic functions have considerable importance for the theory of quadratic factoring. Those that we find especially salient are developed in Appendix C and summarized here.

(Note: These definitions and their consequences are relative to some fixed population over which the variables at issue have a joint frequency or probability distribution as required to define moments and functional dependencies.)

Definitions

Let $\underline{X} = \langle \underline{x}_1, \ldots, \underline{x}_n \rangle$ be any tuple (algebraically, a column vector) of variables. Then a variable $\underline{z}$ is a quadratic function of $\underline{X}$ just in case, for some $\underline{n} \times \underline{n}$ symmetric real matrix $\mathbf{Q}$, $\underline{z} = \underline{X}'\mathbf{Q}\underline{X}$.

The quadratic space, $\mathcal{Q}_X$, generated by variables $\underline{X}$ is the set of all variables that are quadratic functions of $\underline{X}$.

The linear space, $\mathcal{L}_X$, of variables $\underline{X}$ is the space linearly spanned by $\underline{X}$. That is, $\mathcal{L}_X$ comprises all homogeneous linear functions of $\underline{X}$.

A tuple $\underline{X}$ of variables is (implicitly) complete iff $\mathcal{L}_X$ contains unit variable $\underline{x}_0$, and is m(anifestly)-complete iff $\underline{x}_0$ is a component of $\underline{X}$. If $\underline{X}$ is not m-complete, its m-completion is $\langle \underline{x}_0, \underline{X} \rangle$.

Consequences

If $\underline{X}$ is complete, the linear space $\mathcal{L}_X$ of $\underline{X}$ is included in its quadratic space $\mathcal{Q}_X$. That is, the quadratic functions of a complete $\underline{X}$ admit linear terms and additive constants.

If $\underline{X}$ and $\underline{Z}$ linearly span the same space $\mathcal{L}_X = \mathcal{L}_Z$, then $\underline{X}$ and $\underline{Z}$ also generate the same quadratic space $\mathcal{Q}_X = \mathcal{Q}_Z$.

The quadratic space $\mathcal{Q}_X$ generated by variables $\underline{X}$ is also a linear space spanned, inter alia, by $\underline{X}^*$ and by $\underline{X}^{\circledast}$. However, $\mathcal{Q}_X$ is also linearly spanned by many other tuples of quadratic functions of $\underline{X}$ which are not in general quadratic developments of any tuples in $\mathcal{L}_X$.

Hence when we seek to fit quad-factor model (9) to the quad-moments of our data-variables' true-parts $\underline{T} = \langle \underline{t}_1,\ldots,\underline{t}_n \rangle$, although the latter can routinely be decomposed in classic form $\mathbf{M}(\underline{T}^{\otimes},\underline{T}^{\otimes}) = \mathbf{B}\mathbf{M}(\underline{G},\underline{G})\mathbf{B}'$ for one or another linear basis $\underline{G}$ of $\mathcal{Q}_T$, an arbitrary choice of 2nd-level factors $\underline{G}$ will almost certainly __not__ be the quadratic development of any 1st-level factor basis for $\mathcal{L}_T$. This raises quad-factoring's __alignment problem__: When decomposing the 1st- and 2nd-level true-part moments jointly as $\mathbf{M}(\underline{T},\underline{T}) = \mathbf{A}\mathbf{M}(\underline{F},\underline{F})\mathbf{A}'$ and $\mathbf{M}(\underline{T}^{\otimes},\underline{T}^{\otimes}) = \mathbf{B}\mathbf{M}(\underline{G},\underline{G})\mathbf{B}'$, how do we contrive further to have $\underline{G} = \underline{T}^{\otimes}$ or at least $\mathbf{M}(\underline{G},\underline{G}) = \mathbf{M}(\underline{F}^{\otimes},\underline{F}^{\otimes})$? As the Uniqueness Theorem, below, will show, the answer is happily straightforward.

> If $\underline{X}$ is a basis for its linear space $\mathcal{L}_X$, $\underline{X}^*$ fails to be a linear basis for $\mathcal{Q}_X$ just in case, for some tuple $\underline{Z}$ of variables in $\mathcal{L}_X$, all joint scores on $\underline{Z}$ lie on a hyperbolic surface.

The significance of this theorem is, first of all, that $\mathbf{M}(\underline{X}^*,\underline{X}^*)$ can be singular even when $\mathbf{M}(\underline{X},\underline{X})$ is not, and secondly that singular $\mathbf{M}(\underline{X}^*,\underline{X}^*)$ can arise in ways other than the one that seems most interpretively significant when $\underline{X}^*$ is $\underline{m}$-complete (see. p. 25ff., below).

__Tensor-algebraic formulations of quad-factoring relations.__

For any tuple $\underline{X}$ of variables, the full quadratic development $\underline{X}^{\otimes}$ of $\underline{X}$ can be written as the Kronecker product of $\underline{X}$ with itself. That is,

$$\underline{X}^{\otimes} =_{\text{def}} \underline{\text{vec}}(\underline{X}\underline{X}') = \underline{X} \otimes \underline{X} \ .$$

If $\underline{Z}$ is in the linear space $\mathcal{L}_X$ of $\underline{X}$, so that $\underline{Z} = \mathbf{A}\underline{X}$ for some coefficient matrix A, $\underline{X}^{\otimes}$ determines $\underline{Z}^{\otimes}$ according to

$$\underline{Z}^{\otimes} = \underline{Z} \otimes \underline{Z} = \mathbf{A}\underline{X} \otimes \mathbf{A}\underline{X} = (\mathbf{A} \otimes \mathbf{A})(\underline{X} \otimes \underline{X}) = (\mathbf{A} \otimes \mathbf{A})\underline{X}^{\otimes} \ .$$

Moreover, if $\underline{X}$ is a basis for $\mathcal{L}_X$, 1st-level coefficient matrix $\mathbf{A}$ has a left-inverse $\mathbf{A}^L$ such that $\mathbf{A}^L\mathbf{A} = \mathbf{I}$, whence $\underline{X}$ and $\underline{X}^{\otimes}$ can be recovered from $\underline{Z}$ by

$$X = A^L Z , \qquad X^{\circledast} = (A^L \circledast A^L) Z^{\circledast} \quad .$$

Evidently we have not merely

$$M(Z,Z) = A M(X,X) A' , \qquad M(X,X) = A^L M(Z,Z) A^{L\prime} \quad ,$$

in this case but also

$$M(Z^{\circledast},Z^{\circledast}) = (A \circledast A) M(X^{\circledast},X^{\circledast}) (A \circledast A)' , \quad M(X^{\circledast},X^{\circledast}) = (A^L \circledast A^L) M(Z^{\circledast},Z^{\circledast}) (A^L \circledast A^L)' \; .$$

<u>The quad-factoring uniqueness theorem</u>. Suppose that the quad-moments of variables X have a decomposition of form

$$M(X^{\circledast},X^{\circledast}) = (A \circledast A) M_g (A \circledast A)'$$

for some identified matrix A having a left-inverse A^L. Then there exists a tuple of 1st-level factors F of X, namely $F =_{def} A^L X$, such that

$$X = AF , \qquad X^{\circledast} = (A \circledast A) F^{\circledast} , \qquad M(F,F) = M_g \; .$$

Moreover, for any tuple of variables G in $\mathcal{Q}_X$ that reproduces the quad-moments of Y by this same 2nd-level pattern $A \circledast A$, i.e. for which $M(X^{\circledast},X^{\circledast}) = (A \circledast A) M(G,G) (A \circledast A)'$, we have $G = F_a^{\circledast}$ for some 1st-level factor tuple F_a only if F_a differs from F by at most a reflection of axes.

Hence we solve the alignment problem by imposing the constraint that the pattern matrix in our decomposition of true-part quad-moments $M(T^{\circledast},T^{\circledast})$ have structure $A \circledast A$ for a left-invertible 1st-level pattern matrix A. Choice of A is non-unique in the very same way that 1st-level factor patterns are nonunique. But whatever side conditions suffice to select a specific A in $M(T,T) = AM(F,F)A'$ (notably, accounted-for-variance maximization for initial extraction, eventually followed by rotation to simple structure) also suffice to identify a factor tuple satisfying the quad-factoring model that is essentially unique relative to T and A.

Inductive solutions for 2nd-level factor patterns.

To embed 1st-level factoring of data variables Y in 2nd-level factoring of $Y^{@}$ (or rather, in practice, of Y^*), we must include the unit variable among the 1st-level factors as well as, for conceptual convenience, among the variables factored. Accordingly, we expand the orthodox 1st-level arrays of data variables $Y = \langle y_1,\ldots,y_n \rangle$, their true-parts $T = \langle t_1,\ldots,t_n \rangle$, and their common factors $F = \langle f_1,\ldots,f_r \rangle$ into their respective m-completions $Y_0 = \langle y_0, Y \rangle$, $T_0 = \langle t_0, T \rangle$, and $F_0 = \langle f_0, F \rangle$. (Reminder: y_0, t_0, and f_0 are all constant at unity.) Then augmenting (1) by the trivial $y_0 = t_0 + e_0$ (e_0 constant at zero) extends our 1st-level data variables' true-part/error decomposition to

$$Y_0 = T_0 + E_0 \tag{10}$$

while 1st-level factor model (7) becomes

$$t_i = \sum_{j=0}^{r} a_{ij} f_j \qquad (i = 0,1,\ldots,n)$$

or equivalently

$$T_0 = \mathbf{A}F_0 \tag{11}$$

wherein $\mathbf{A}$ is of course the $(1+n) \times (1+r)$ matrix of pattern coefficients $\{a_{ij}\}$.

Compared to orthodox 1st-level factor models, pattern matrix $\mathbf{A}$ in (11) has an extra row and an extra column. Its extra row, the pattern for t_0, is inflexibly all zero except $a_{00} = 1$. In contrast, the added first column of $\mathbf{A}$, i.e. the 1st-level pattern coefficients $\{a_{i0}\}$ on unit factor f_0, is open to a variety of numerical specifications. Whether these make (11) differ more than trivially from conventional factoring depends on whether F is constrained by orthogonality to f_0. (We use "orthogonality" here in its generic sense of zero expected pairwise products rather than its special sense of zero covariances.) If $f_1,\ldots,f_r$ are required as usual to have zero means, i.e. to be orthogonal to f_0, then $a_{i0} = \mathcal{E}[t_i] = m_{y_i}$ for each $i = 1,\ldots,n$--whence under centered scaling for Y the first column of $\mathbf{A}$ becomes all zero save $a_{00} = 1$. But allowing 1st-level variables Y to retain natural means has no effect in this case on the rest of $\underset{\sim}{A}$. That is, so long as $\underline{F}$ is orthogonal to $\underline{f}_0$, the part of $\underset{\sim}{A}$ that remains after deletion of its first row and column is some more-or-less orthodox pattern obtainable by factoring the $\underline{Y}$-covariances without regard for how the $\underline{Y}$-means are scaled.

On the other hand, if factors $\underline{f}_1,\ldots,\underline{f}_r$ in (11) are <u>not</u> all orthogonal to $\underline{f}_0$, each a_{i0} continues to be the additive constant in $\underline{y}_i$'s regression upon $\underline{F}$ but almost certainly differs from m_{y_i}. Allowing the $\underline{F}$-means to be nonzero is not only unconventional but would usually be unmotivated as well, especially for 1st-level factoring of centered data. Yet there do exist circumstances of quad-factoring, and even occasionally of ordinary 1st-level factoring, in which it makes interpretive sense to allow factor rotations in which $\underline{F}$ becomes oblique to $\underline{f}_0$. Quadratic factors are best initially extracted under orthogonality of $\underline{F}$ to $\underline{f}_0$; but eventually we may find reasons to relax this constraint.

(Once we consider rotation of (11), still another possibility for the extended 1st-level pattern is to let this comprise the coefficients for $\underline{T}_0$ on some basis $\underline{F}_1$ for $\underline{F}_0$-space in which rotated axis tuple $\underline{F}_1 = \underset{\sim}{W}\underline{F}_0$ is not m-complete. But we can think of no meaningful interpretation for the pattern $\underset{\sim}{A}\underset{\sim}{W}^{-1}$ on factors so positioned.)

Because $\underline{Y}_0 = \langle \underline{y}_0, \underline{Y}\rangle = \langle \underline{y}_0, \underline{y}_1,\ldots,\underline{y}_n\rangle$ is m-complete, its full quadratic development

$$\underline{Y}_0^{\circledast} =_{def} \underline{\text{vec}}(\underline{Y}_0\underline{Y}_0') = \underline{Y}_0 \circledast \underline{Y}_0$$

comprises not merely the proper 2nd-level product variables $\{\underline{y}_i\underline{y}_j : \underline{i},\underline{j} = 1,\ldots,\underline{n}\}$ but all 1st-level data variables $\{\underline{y}_i\ (= \underline{y}_0\underline{y}_i) : \underline{i} = 1,\ldots,\underline{n}\}$ and unit variable $\underline{y}_0$ $(= \underline{y}_0\underline{y}_0)$ as well. The true-part/error decomposition of $\underline{Y}_0^{\circledast}$ is of course

$$\begin{aligned}\underline{Y}_0^{\circledast} &= (\underline{T}_0 + \underline{E}_0) \circledast (\underline{T}_0 + \underline{E}_0) = (\underline{T}_0 \circledast \underline{T}_0) + (\underline{T}_0 \circledast \underline{E}_0) + (\underline{E}_0 \circledast \underline{T}_0) + (\underline{E}_0 \circledast \underline{E}_0)\\ &= \underline{T}_0^{\circledast} + \underline{E}_0^{\circledast}\end{aligned}$$

wherein $\underline{T}_0^{\circledast} = \underline{T}_0 \circledast \underline{T}_0$ is the true-part of $\underline{Y}_0$'s full quadratic development $\underline{Y}_0^{\circledast}$ while the residual $\underline{E}_0^{\oplus}$ thereof is

$$\underline{E}_0^{\oplus} =_{\text{def}} \underline{Y}_0^{\otimes} - \underline{T}_0^{\otimes} = (\underline{T}_0 \otimes \underline{E}_0) + (\underline{E}_0 \otimes \underline{T}_0) + \underline{E}_0^{\otimes} \quad .$$

So the 2rd-order moment matrix for $\underline{Y}_0^{\otimes}$--which by virtue of $\underline{Y}_0$'s m-completeness actually comprises all $\underline{Y}$-moments through the 4th order--has composition

$$\begin{aligned} \underset{\sim}{M}(\underline{Y}_0^{\otimes},\underline{Y}_0^{\otimes}) &= \underset{\sim}{M}(\underline{T}_0^{\otimes},\underline{T}_0^{\otimes}) + \underset{\sim}{M}(\underline{T}_0^{\otimes},\underline{E}_0^{\oplus}) + \underset{\sim}{M}(\underline{E}_0^{\oplus},\underline{T}_0^{\otimes}) + \underset{\sim}{M}(\underline{E}_0^{\oplus},\underline{E}_0^{\oplus}) \\ &= \underset{\sim}{M}(\underline{T}_0^{\otimes},\underline{T}_0^{\otimes}) + \underset{\sim}{Q}(\underline{E}_0^{\oplus}) \quad , \end{aligned} \tag{12}$$

where total-error matrix $\underset{\sim}{Q}(\underline{E}_0^{\oplus})$ (see definition (5)) is specified by $\underset{\sim}{M}(\underline{Y},\underline{Y})$ and the uniqueness parameters--namely $\underset{\sim}{u} = \langle \underline{u}_1,\ldots,\underline{u}_n \rangle$ and, if not presumed Normal, $\underset{\sim}{u}^{[3]} = \langle \underline{u}_1^{[3]},\ldots,\underline{u}_n^{[3]} \rangle$ and $\underset{\sim}{u}^{[4]} = \langle \underline{u}_1^{[4]},\ldots,\underline{u}_n^{[4]} \rangle$--according to (6). Conditional on our choice of error-model strength, let us say

$$\underset{\sim}{u}^{+} =_{\text{def}} \begin{cases} \underset{\sim}{u} & \text{if Normality of both } \underset{\sim}{u}^{[3]} \text{ and } \underset{\sim}{u}^{[4]} \text{ is presumed} \\ \langle \underset{\sim}{u},\underset{\sim}{u}^{[4]} \rangle & \text{if Normality just of } \underset{\sim}{u}^{[3]} \text{ is presumed} \\ \langle \underset{\sim}{u},\underset{\sim}{u}^{[4]},\underset{\sim}{u}^{[3]} \rangle & \text{if no error Normalities are presumed .} \end{cases}$$

(These are the only a priori error-model alternatives that we have programmed. But additional variants would be routine to include were not need for them obviated by our new technique, described in Appendix B, for ad hoc relaxation of (6) at points of greatest model misfit.) It is straightforward to program specifications (6) into an algorithm that maps $\underset{\sim}{M}(\underline{Y}_0,\underline{Y}_0)$ and any numerical estimate of $\underset{\sim}{u}^{+}$ into a corresponding numerical estimate of $\underset{\sim}{Q}(\underline{E}_0^{\oplus})$ and from there of $\underset{\sim}{M}(\underline{T}_0^{\otimes},\underline{T}_0^{\otimes})$. And starting from any initial estimate of $\underset{\sim}{u}^{+}$ (as provided, say, by orthodox 1st-level factoring of $\underset{\sim}{C}(\underline{Y},\underline{Y})$ along with the strong error model), we are able to iterate improvements on this as the analysis progresses. So estimating $\underset{\sim}{M}(\underline{T}_0^{\otimes},\underline{T}_0^{\otimes})$ is essentially routine. Our main problem is how to convert the latter, in turn, into richer information about factors $\underline{F}$ and their determination of $\underline{Y}$ than can be extracted just from $\underset{\sim}{C}(\underline{Y},\underline{Y})$.

According to model (11), the full quadratic development of our data variables' true-parts has composition

$$\underline{T}_0^{@} = \underline{T}_0 \circledast \underline{T}_0 = \underset{\sim}{A}\underline{F}_0 \circledast \underset{\sim}{A}\underline{F}_0 = (\underset{\sim}{A} \circledast \underset{\sim}{A})(\underline{F}_0 \circledast \underline{F}_0) = (\underset{\sim}{A} \circledast \underset{\sim}{A})\underline{F}_0^{@} \tag{13}$$

So the quad-moments of $\underline{T}_0$ decompose as

$$\underset{\sim}{M}(\underline{T}_0^{@},\underline{T}_0^{@}) = (\underset{\sim}{A} \circledast \underset{\sim}{A})\ \underset{\sim}{M}(\underline{F}_0^{@},\underline{F}_0^{@})\ (\underset{\sim}{A} \circledast \underset{\sim}{A}) \quad , \tag{14}$$

and the task of quadratic factor analysis is to find estimates of $\underset{\sim}{A}$ and $\underset{\sim}{M}(\underline{F}_0^{@},\underline{F}_0^{@})$ which, together with our estimates of the error terms in $\underset{\sim}{Q}(\underline{E}_0^{@})$, tidily reproduce data quad-moment matrix $\underset{\sim}{M}(\underline{Y}_0^{@},\underline{Y}_0^{@})$. Or rather, this is quad-factoring's theoretically perspicuous description. In practice, since (13) relates the full quadratic development of $\underline{T}_0$ to that of $\underline{F}_0$, there are massive redundancies in (14) that make direct analysis of $\underset{\sim}{M}(\underline{T}_0^{@},\underline{T}_0^{@})$ inexpedient. Far easier is to work instead with the counterparts of (13/14) for the <u>bare</u> quadratic developments of $\underline{T}_0$ and $\underline{F}_0$, namely,

$$\underline{T}_0^{*} = \underset{\sim}{A}_{*}\underline{F}_0^{*} \quad , \tag{13a}$$

$$\underset{\sim}{M}(\underline{T}_0^{*},\underline{T}_0^{*}) = \underset{\sim}{A}_{*}\ \underset{\sim}{M}(\underline{F}_0^{*},\underline{F}_0^{*})\ \underset{\sim}{A}_{*}' \quad , \tag{14a}$$

in which the elements of $\underset{\sim}{A}_{*}$ are derived from those of $\underset{\sim}{A}$ according to formula (9) expanded to include index 0. That is, for $\underline{h} = 0,1,\ldots,\underline{n}$, $\underline{i} = \underline{h},\ldots,\underline{n}$, $\underline{j} = 0,1,\ldots,\underline{r}$, $\underline{k} = \underline{j},\ldots,\underline{r}$,

$$[\underset{\sim}{A}_{*}]_{hi,jk} = \begin{cases} [\underset{\sim}{A}]_{hj}[\underset{\sim}{A}]_{ik} + [\underset{\sim}{A}]_{hk}[\underset{\sim}{A}]_{ij} & \text{if } \underline{j} < \underline{k} \\ [\underset{\sim}{A}]_{hj}[\underset{\sim}{A}]_{ik} & \text{if } \underline{j} = \underline{k} \end{cases} \quad \left([\underset{\sim}{A}]_{0i} = \begin{cases} 1 \text{ if } \underline{i} = 0 \\ 0 \text{ if } \underline{i} > 0 \end{cases} \right). \tag{15}$$

Since $\underset{\sim}{A}$ is the upper-left $(1+\underline{n}) \times (1+\underline{r})$ submatrix of both $\underset{\sim}{A} \circledast \underset{\sim}{A}$ and $\underset{\sim}{A}_{*}$, any one of $\{\underset{\sim}{A},\ \underset{\sim}{A}_{*},\ \underset{\sim}{A} \circledast \underset{\sim}{A}\}$ identifies the other two. And $\underset{\sim}{M}(\underline{T}_0^{@},\underline{T}_0^{@})$ strips down to $\underset{\sim}{M}(\underline{T}_0^{*},\underline{T}_0^{*})$ by deleting from the former all rows $\underline{hi}$ and columns $\underline{jk}$ for which $\underline{h} > \underline{i}$ or $\underline{j} > \underline{k}$. Operationally, we disregard $\underset{\sim}{M}(\underline{T}_0^{@},\underline{T}_0^{@})$ altogether and instead estimate $\underset{\sim}{M}(\underline{T}^{*},\underline{T}^{*})$ directly from $\underset{\sim}{M}(\underline{Y}^{*},\underline{Y}^{*})$ and our running estimate of the error terms in (12)'s counterpart

$$\begin{aligned} \underset{\sim}{M}(\underline{Y}^{*},\underline{Y}^{*}) &= \underset{\sim}{M}(\underline{T}_0^{*},\underline{T}_0^{*}) + \underset{\sim}{M}(\underline{T}_0^{*},\underline{E}_0^{+}) + \underset{\sim}{M}(\underline{E}_0^{+},\underline{T}_0^{*}) + \underset{\sim}{M}(\underline{E}_0^{+},\underline{E}_0^{+}) \\ &= \underset{\sim}{M}(\underline{T}_0^{*},\underline{T}_0^{*}) + \underset{\sim}{Q}(\underline{E}_0^{+}) \quad . \end{aligned} \tag{12a}$$

(We have previously written (12a) as equations (4) and (4').) Combining our two moment models--the one for errors and the one for factors--into a single equation, we can then say that quad-factoring is a decomposition of the data variables' quad-moment matrix having form

$$\underset{\sim}{M}(\underline{Y}_0^{\circledast},\underline{Y}_0^{\circledast}) \;=\; (\underset{\sim}{A} \circledast \underset{\sim}{A})\ \underset{\sim}{M}(\underline{F}_0^{\circledast},\underline{F}_0^{\circledast})\ (\underset{\sim}{A} \circledast \underset{\sim}{A})' + \underset{\sim}{Q}(\underline{E}_0^{\oplus}) \qquad (16)$$

or less redundantly

$$\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*) \;=\; \underset{\sim}{A}_*\ \underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)\ \underset{\sim}{A}_*' \;+\; \underset{\sim}{Q}(\underline{E}_0^+) \quad , \qquad (16a)$$

wherein $\underset{\sim}{A}_*$ has structure (15) while $\underset{\sim}{Q}(\underline{E}_0^{\oplus})$ and its less redundant subarray $\underset{\sim}{Q}(\underline{E}_0^+)$ are specified from $\underset{\sim}{M}(\underline{Y},\underline{Y})$ and uniqueness parameters $\underset{\sim}{u}^+$ by model (6).

In principle, it should be routine to solve the quad-factoring model by any modern structural-modelling logic such as LISREL or RAM (McArdle & McDonald, 1984). The composition of equations (6,15) into equation (16a) defines a computable function Φ from guesses $\langle\hat{\underset{\sim}{u}}^+,\hat{\underset{\sim}{A}},\hat{\underset{\sim}{M}}_F\rangle$ at $\langle\underset{\sim}{u}^+,\underset{\sim}{A},\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)\rangle$ into reproductions of data array $\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*)$. So relative to any chosen loss function, the best estimate of our empirical quad-moments' source parameters is the $\langle\hat{\underset{\sim}{u}}^+,\hat{\underset{\sim}{A}},\hat{\underset{\sim}{M}}_F\rangle$ for which the loss of approximating $\underset{\sim}{M}(\underline{Y}_0^*,\underline{Y}_0^*)$ by $\Phi(\underset{\sim}{u}^+,\underset{\sim}{A},\underset{\sim}{M}_F)$ is minimal. In practice, however, the problem size for quad-factoring even modestly many data variables is so large that we have not yet managed to set up the subroutines required for a complete structural-modelling solution. We have, however, operationalized solutions using more classical routines that allow quad-factoring to be tested in practice even as we seek more powerful algorithms that lessen certain admitted suboptimalities in our present procedure.

In fact, we have devised a spectrum of quad-factoring alternatives, selected by control-parameter specification in our generic QUADFAC program and differing inter alia in how strong an error-model is presumed crossed with what portion of complete residual array (6) is used to estimate $\underset{\sim}{u}$. (QUADFAC 's FORTRAN-77 source code, together with a package of supporting programs, is available. Ask and ye shall receive.) At one extreme--call this "fast QUADFAC"--the routine is computationally quite frugal, albeit by deriving the factor pattern just from the 1st-level data covariances and thus losing the higher-moment pattern information whose exploitation is one of quad-factoring's hoped-for benefits. In contrast, QUADFAC's other versions use all the data quad-moments for identifying the factor pattern, though at computer costs several times that of fast QUADFAC and still not as thoroughly as we hope ultimately to attain. Details of QUADFAC's solution logic are developed in Appendix B, while Appendix D compares QUADFAC's accuracy at parameter recovery from artificial data under all its main procedural variants crossed with variation in factor structure and sampling noise.

Interpretation of results.

Once QUADFAC iteration has converged upon estimates of $\mathbf{u}^{+}$ and the $\langle \mathbf{A}, \mathbf{M}(\underline{Y}_0^*, \underline{Y}_0^*) \rangle$ defined by principal-axes positioning of $\underline{F}_0$ with $\mathbf{M}(\underline{Y}_0^*, \underline{Y}_0^*)$-reproduction loss small enough to warrant taking the results seriously, we turn to final adjustments that enhance meaningfulness of results. (We shall not here distinguish notationally between model parameters and our computed estimates thereof.) First comes rotation of 1st-level factor axes to positions that seemingly make the greatest interpretive sense. Quadratic factor theory is entirely open to any criterion for this; but we shall presume that you share our preference for oblique simple structure.

Rotation of axes.

If 1st-level factor axes $\underline{F}_0$ in $\underline{T}_0 = \mathbf{A}\underline{F}_0$ are rotated to $\underline{G}_0 = \mathbf{W}\underline{F}_0$, the effect thereof on factor pattern at both 1st and 2nd levels is

$$\underline{T}_0 = (\mathbf{A}\mathbf{W}^{-1})\underline{G}_0 \,, \quad \underline{T}_0^* = (\mathbf{A}\mathbf{W}^{-1})_*\underline{G}_0^* \,, \quad \underline{T}_0^{\circledast} = (\mathbf{A}\mathbf{W}^{-1} \circledast \mathbf{A}\mathbf{W}^{-1})\underline{G}_0^{\circledast} = (\mathbf{A} \circledast \mathbf{A})(\mathbf{W} \circledast \mathbf{W})^{-1}\underline{G}_0^{\circledast} \,,$$

where $(\ )_*$ is the function defined by equation (15). And the rotated factor quad-moments are

$$\mathbf{M}(\underline{G}_0^*, \underline{G}_0^*) = \mathbf{W}_*\mathbf{M}(\underline{F}_0^*, \underline{F}_0^*)\mathbf{W}_*' \,, \quad \mathbf{M}(\underline{G}_0^{\circledast}, \underline{G}_0^{\circledast}) = (\mathbf{W} \circledast \mathbf{W})\, \mathbf{M}(\underline{F}_0^{\circledast}, \underline{F}_0^{\circledast})\, (\mathbf{W} \circledast \mathbf{W})' \,.$$

It is evident here that when positioning factor axes, quad-factoring is not limited to selection of $\mathbf{W}$ just in light of what this does to rotated 1st-level pattern $\mathbf{A}\mathbf{W}^{-1}$

but can examine its effect on the much larger coefficient array $(A \otimes A)(W \otimes W)^{-1}$. We might hope, therefore, that simple-structure hyperplanes can be discerned more sharply in a quadratic factor pattern than are clear in just the embedded 1st-level pattern. And to our surprise we find that solving for W in $(A \otimes A)(W \otimes W)^{-1}$ to maximize 2nd-level hyperplane strength is indeed operationally feasible. Disappointingly, however, the theory of this shows also that 2nd-level rotation of the pattern in $T_0^{\otimes} = (A \otimes A)F_0^{\otimes}$ is virtually equivalent to rotating the 1st-level pattern in

$$\begin{bmatrix} \vdots \\ a_{ij}T_0 \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ a_{ij}A \\ \vdots \end{bmatrix} F_0$$

for the aggregate of all different rescalings $\{(a_{ij}T_0) = (a_{ij}A)F_0\}$ of $T_0 = AF_0$ by the various elements a_{ij} of A. And there is no evident reason why any such aggregated multicopying of 1st-level pattern A should demark hyperplanes more clearly than does A by itself. (If you look at the multicopied pattern plot for one pair of factor axes, you'll see what we mean.)

Accordingly, with one important exception (namely, cases where we suspect that some dimensions of $\mathcal{L}_{F_0}$ are quad-functions of others--see below), we recommend that factor axes be terminally positioned by rotating just the 1st-level part A of initial 2nd-level pattern A_* to simple structure by whatever algorithm for this you prefer, with subsequent use of the W so found to compute the rotated factor quad-moments (and, if you want it, the rotated 2nd-level pattern) as shown above. (If you feed your QUADFAC output into the HYBALL program for 1st-level factor rotation described in Rozeboom, 198 , your rotated pattern printout will be automatically accompanied by the rotated factor quad-moments.) And we also recommend constraining this rotation to form

$$\begin{bmatrix} g_0 \\ G \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & W_F \end{bmatrix} \begin{bmatrix} f_0 \\ F \end{bmatrix}$$

with rotation of initial pattern $\underset{\sim}{A} = \begin{bmatrix} 1 & \underset{\sim}{0} \\ \underset{\sim}{m}_Y & \underset{\sim}{A}_F \end{bmatrix}$ correspondingly restricted to form

$$\underset{\sim}{A}\underset{\sim}{W}^{-1} = \begin{bmatrix} 1 & \underset{\sim}{0} \\ \underset{\sim}{m}_Y & \underset{\sim}{A}_F\underset{\sim}{W}_F^{-1} \end{bmatrix} .$$

This keeps the rotation just within the subspace of $\underline{F}_0$ orthogonal to $\underline{f}_0$. As observed earlier (p. 17), the main alternative to this constraint is to fix $\underline{f}_0$ (= $\underline{g}_0$) but allow $\underline{G}$ to become oblique to $\underline{g}_0$. A minor reprogramming of HYBALL can easily accomplish this, but it serves no purpose unless data variables $\underline{Y}$ have non-arbitrary means. For allowing obliquity of $\underline{G}$ to $\underline{g}_0$ affects the pattern attainable on $\underline{G}$ only in the column scalings that normalize factor variances; and although it can simplify the pattern on $\underline{f}_0$ when this initially contains natural means, the first column of $\underset{\sim}{A}$ is already ideal by artifice when the data variables are centered.

What to do with factor quad-moments.

Let us revert to notation "$\underline{F}$" for the 1st-level factors we hope to interpret, however these may have been repositioned after initial extraction. Now that our solution for $\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)$ has given us the $\underline{F}$-distribution's moments through the 4th order, what good is this information?

Having raised this question, we must confess that our ability to answer it is still rather limited. But the obvious first interpretive step is to check out $\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)$'s compatability with our sample $\underline{F}$-distribution's being viewed as approximately Normal. Were $\underline{F} = \langle \underline{f}_1,\ldots,\underline{f}_r \rangle$ to be Normally distributed, with the 1st-level moment matrix for its m-completion Gram-factorable as $\underset{\sim}{M}(\underline{F}_0,\underline{F}_0) = \underset{\sim}{W}\underset{\sim}{W}'$, the bare quad-moment matrix for $\underline{F}_0 = \langle \underline{f}_0,\underline{F} \rangle$ would be

$$\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*) = \underset{\sim}{W}_*\underset{\sim}{K}\underset{\sim}{W}_*'$$

wherein $\underset{\sim}{K}$ is the bare quad-moment matrix for the m-completion of any $\underline{r}$-tuple of Normal variables that are also centered and orthonormal. Specifically,

$$[\underset{\sim}{K}]_{ii,jj} = [\underset{\sim}{K}]_{jj,ii} = [\underset{\sim}{K}]_{ij,ij} = 1 \quad (\underline{i} < \underline{j})$$

$$[\underset{\sim}{K}]_{ii,ii} = \begin{cases} 1 & \text{if } \underline{i} = 0 \\ 3 & \text{if } \underline{i} > 0 \end{cases}$$

$$[K]_{hi,jk} = 0 \text{ otherwise} .$$

So computing W from the 1st-level part of $M(F_0^*,F_0^*)$ (or simply taking $W = I$ if F_0 is our initial orthonormal solution) and comparing $W_*^{-1}M(F_0^*,F_0^*)W_*^{-1\prime}$ (or simply $M(F_0^*,F_0^*)$ for orthonormal F_0) to K appraises the degree of Normality in F's quad-moments. If this comparison discredits the hypothesis of factor Normality (a judgment which by rights should include some statistical testing whose analytic development lies beyond our competence), whatever features of the rotated factor quad-moments appear most saliently nonNormal stand as empirical disclosures awaiting explaination by substantive theories of these data.

Generic interpretation of nonNormality in factor quad-moments is still largely <u>terra incognita</u> for us. Even so, we direct your attention to two special prospects, one minor but the other major. The first is diagnosis whether any of the F-factors are dichotomous. Despite the optimism of Gangestad & Snyder (1985), however, we doubt that many dichotomous source variables are out there awaiting detection. More provocative is the prospect that arises when near-zero roots in $M(F_0^*,F_0^*)$ reveal multicollinearities among factors $F_0^* = \langle f_0, F, F^* \rangle$. Whether this has any generic significance deeper than the hyperbolic-surface theorem reported on p. 14, above, we do not know. But one outstandingly important way for F_0^* to contain linear dependencies is for some of 1st-level factors $F_0 = \langle f_0, f_1, \ldots, f_r \rangle$ to be quadratic functions of the others. For f_i is in the quadratic space of, say, $X_0 = \langle f_0, f_1, \ldots, f_s \rangle$ $(s < r)$ just in case it is in the linear space of F_0^*'s subtuple X_0^*. And if $f_{s+1}, \ldots, f_r$ are all quadratic functions of X_0, then the 1st-level data variables' true parts that we have found to be linearly decomposable as $T_0 = AF_0$ are really quad-functions just of X_0. So quad-factoring is in effect also a version of nonlinear factor analysis (see McDonald, 1967; Etezadi-Amoli & McDonald, 1983) --not however by coersion but by permissive discovery.

<u>Diagnosis of dichotomies</u>. For any variable x with mean μ_x and variance σ_x^2, the skew sk_x and kurtosis kt_x in a given distribution of x may be defined

$$\underline{sk}_x =_{def} \mathcal{E}[(\underline{x} - \underline{m}_x)^3]/\sigma_x^3 , \quad \underline{kt}_x =_{def} \mathcal{E}[(\underline{x} - \underline{m}_x)^4]/\sigma_x^4 .$$

(We depart here from the tradition of defining kurtosis as kt minus 3. The subtraction makes a comparison to Normality that analytically is a useless complication.) And for the 1st-level factors $\underline{F} = \langle \underline{f}_1, \ldots, \underline{f}_r \rangle$ whose quad-moments are found by QUADFAC under assignment of standard scaling, this becomes simply

$$\underline{sk}_{f_i} = [\underset{\sim}{M}(\underline{F}_0^*, \underline{F}_0^*)]_{0i,ii} , \quad \underline{kt}_{f_i} = [\underset{\sim}{M}(\underline{F}_0^*, \underline{F}_0^*)]_{ii,ii} .$$

Now, it is easy to show that if numerically scaled variable $\underline{x}$ is dichotomous, with $\underline{p}_x$ ($\underline{q}_x$) the population proportion in its higher (lower) category,

$$\underline{kt}_x + 3 = (\underline{p}_x\underline{q}_x)^{-1} = \underline{sk}_x^2 + 4 \quad (\text{ dichotomous } \underline{x}) .$$

So quad-factoring appraises whether 1st-level factor $\underline{f}_i$ is dichotomous by judging whether $[\underset{\sim}{M}(\underline{F}_0^*, \underline{F}_0^*)]_{ii,ii}$ is essentially equal to $1 + [\underset{\sim}{M}(\underline{F}^*, \underline{F}^*)]_{0i,ii}^2$. Unhappily, our performance studies show that with noisy data, QUADFAC's present computations often overestimate factor kurtosis, sometimes disagreeably so. But we are confident that reliability of the factor quad-moment solution can be substantially improved.

Diagnosis of quadratic factor dependencies. In principle, it is entirely straightforward to determine which dimensions of $\underline{F}$-space, if any, are quadratic functions of others. Suppose that $\underline{Z}$ and $\underline{X}$ are subsets of factors $\underline{F} = \langle \underline{f}_1, \ldots, \underline{f}_r \rangle$, or of some rotation of $\underline{F}$, while $\underline{X}_0 = \langle \underline{x}_0, \underline{X} \rangle$ is the m-completion of $\underline{X}$. ($\underline{X}$ and $\underline{Z}$ need not be disjoint; in fact, for some purposes we want $\underline{Z} = \underline{F}$.) Then the quadratic regression of $\underline{Z}$ upon $\underline{X}$ is $\dot{\underline{Z}} = \underset{\sim}{B}_Z \underline{X}_0^*$ for coefficient matrix

$$\underset{\sim}{B}_Z = \underset{\sim}{M}(\underline{Z}, \underline{X}_0^*) \ \underset{\sim}{M}^+(\underline{X}_0^*, \underline{X}_0^*) ,$$

where $\underset{\sim}{M}^+(\underline{X}_0^*, \underline{X}_0^*)$ is the inverse or, when necessary, the pseudo-inverse of $\underline{X}$'s quad-moment matrix. And the diagonal of

$$\underset{\sim}{M}(\underline{Z}, \underline{Z}; \underline{X}_0^*) = \underset{\sim}{M}(\underline{Z}, \underline{Z}) - \underset{\sim}{M}(\underline{Z}, \underline{X}_0^*) \ \underset{\sim}{M}^+(\underline{X}_0^*, \underline{X}_0^*) \ \underset{\sim}{M}'(\underline{Z}, \underline{X}_0^*) \qquad (17)$$

comprises the residual variances of factors $\underline{Z}$ after their quadratic regression on $\underline{X}$

is partialled out. When $\underline{X}$ and $\underline{Z}$ are both subtuples of $\underline{F}$, all terms on the right in (17) are contained in $\underset{\sim}{M}(\underline{F}^*,\underline{F}^*)$; whence to judge which factors in $\underline{Z}$ are (nearly) quad-functions of factor subtuple $\underline{X}$ we need only compute the diagonal elements of $\underset{\sim}{M}(\underline{Z},\underline{Z};\underline{X}_0^*)$ and note which ones are (nearly) zero. Suppose that when $\underline{X}_0$ and $\underline{Z}$ jointly span $\mathcal{L}_{F_0}$, all factors in $\underline{Z}$ pass this zero-residuals test. Then all dimensions in linear $\underline{F}_0$-space, the true-parts of $\underline{Y}$ in particular, are quad-functions just of $\underline{X}_0$. And the composition of $\dot{\underline{Z}} = \underset{\sim}{B}_Z\underline{X}_0^*$ into the components of $\underline{Z}$ on the right in $\underline{T}_0 = \underset{\sim}{A}\underline{F}_0$ yields coefficients for the putative quadratic determination of $\underline{T}_0$ and hence $\underline{Y}_0$ by 1st-level factors $\underline{X}_0$.

Practical application of this quad-dependency diagnostic, however, incurs a complication whose management seems clear in theory but requires nonlinear-optimization programming that we have not yet accomplished: What dimensions of $\underline{F}$-space should we pick for $\underline{X}$ and $\underline{Z}$? When rotation of 1st-level axes has properly aligned $\underline{F}$ with genuine causal sources of data variables $\underline{Y}$, it suffices to apply (17) to each partition $\langle\underline{X},\underline{Z}\rangle$ of $\underline{F}$, with the number $\underline{s}$ of dimensions in $\underline{X}$ taken first to be $\underline{s} = \underline{r} - 1$, next to be $\underline{s} = \underline{r} - 2$, and so on, stopping when no $\underline{s}$-selection from $\underline{F}$ quadratically accounts adequately for $\underline{F}$'s remainder. But interpretively optimal factor positioning is a chancy attainment at best. Our only decent criterion for this is simple structure; yet it does not take much meditation on the logic of single-plane rotation to appreciate how unreliable we must expect this to be. And when we suspect that some of the factors in a suitably rotated $\underline{F}$ are quad-functions of others, simple structure is not even appropriate in all planes: When $\underline{f}_j$ is a quad-function of, inter alia, $\underline{f}_i$, we still wish to maximize the number of pattern points in the $\underline{f}_i/\underline{f}_m$ plane that lie close to the $\underline{f}_i$-axis; but there is no rationale for trying to achieve the same for the $\underline{f}_j$-axis unless data variables $\underline{Y}$ all have natural means, i.e., no scale centering. (To appreciate this point, consider the simple structure of Galileo's law of falling bodies before and after centering in a distribution of distance-and-duration-of-travel observations.) And proper axis placement is crucial for disclosure of factor quad-dependencies by diagnostic (17), insomuch as when $\underline{F}$ has an $\langle \underline{X},\underline{Z} \rangle$ partition for which non-null $\underline{Z}$ is quad-dependent on $\underline{X}$, this does not generally remain true under rotation of $\underline{F}$.

For any $\underline{s} < \underline{r}$, one way to find $\underline{s}$ independent dimensions $\underline{X}$ of linear $\underline{F}_0$-space that best fit the hypothesis $\mathcal{L}_{F_0} \subseteq \mathcal{Q}_{X_0}$--the simplest we have been able to envision--is as follows: Starting with $\underline{F}_0$ orthonormal, let $\underset{\sim}{R}$ be an arbitrary $(1+\underline{s}) \times (1+\underline{r})$ row-wise orthonormal coefficient matrix whose first row and column are all zero except a leading 1. Then

$$\underline{X}_0 = \langle \underline{x}_0, \underline{X} \rangle =_{\text{def}} \underset{\sim}{R}\underline{F}_0$$

is an m-complete orthonormal basis for some $(1+\underline{s})$-dimensional subspace of $\mathcal{L}_{F_0}$, while the bare quadratic development of $\underline{X}_0$ is

$$\underline{X}_0^* = \underset{\sim}{R}_* \underline{F}_0^*$$

with $\underset{\sim}{R}_*$ defined from $\underset{\sim}{R}$ by the form-(15) expansion. If each $\underline{F}_0$-factor is in $\mathcal{Q}_{X_0}$, as we hope to achieve by suitable choice of $\underline{s}$ and $\underset{\sim}{R}$, there exists some coefficient matrix $\underset{\sim}{B}_F$ such that $\underline{F}_0 = \underset{\sim}{B}_F \underline{X}_0^* = \underset{\sim}{B}_F \underset{\sim}{R}_* \underline{F}_0^*$; whence

$$\underset{\sim}{M}(\underline{F}_0,\underline{F}_0) = (\underset{\sim}{B}_F \underset{\sim}{R}_*)\, \underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)\, (\underset{\sim}{B}_F \underset{\sim}{R}_*)' \quad . \tag{18}$$

Although we have not yet accomplished the programming, solution of (18) for best-fitting $\underset{\sim}{B}_F$ and $\underset{\sim}{R}$ is a straightforward application of modern structural modelling. Moreover, since $\underset{\sim}{M}(\underline{F}_0,\underline{F}_0^*) = (\underset{\sim}{B}_F \underset{\sim}{R}_*)\underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0^*)$, (18) can be simplified to

$$\underset{\sim}{M}(\underline{F}_0,\underline{F}_0) = (\underset{\sim}{B}_F \underset{\sim}{R}_*)\, \underset{\sim}{M}(\underline{F}_0^*,\underline{F}_0) \quad , \tag{19}$$

albeit we are not sure how easily extant structural-modelling programs can be adapted to (19)'s asymmetry in its unknowns. Once a solution of (18) or (19) is in hand, 1st-level pattern $\underset{\sim}{A}$ of $\underline{Y}_0$ upon initial factors $\underline{F}_0$ converts immediately into coefficient matrix $\underset{\sim}{A}\underset{\sim}{B}_F$ of $\underline{Y}_0$'s quad-dependency upon $\underline{X}_0$ $(= \underset{\sim}{R}\underline{F}_0)$.

For fixed $\underline{s}$, the solution for best-fitting B_F and R in (18) or (19) is unique only under side conditions defining an arbitrary placement of axes in $\underline{X}$-space. So once we have found $\underline{T}_0$'s quadratic determination $\underline{T}_0 = AB_F\underline{X}_0^*$ by the initially positioned $\underline{X}$ we want to search out a transformation matrix W in

$$\underline{T}_0 \quad = \quad AB_F W_*^{-1}(W\underline{X}_0)^* \tag{20}$$

that rotates $\underline{X}$ to an interpretively optimal pattern on $(W\underline{X}_0)^*$. Although we are unable to solve $AB_F W_*^{-1}$ directly for simple structure, it is straightforward to rewrite (20) as $\underline{T}_0$'s linear dependency on the full quadratic development $(W\underline{X}_0)^{\otimes}$ $= W\underline{X}_0 \otimes W\underline{X}_0$ of the rotated $\underline{X}$-axes, namely,

$$\underline{T}_0 \quad = \quad AB_Q(W^{-1} \otimes W^{-1})(W\underline{F}_0)^{\otimes} \tag{21}$$

where B_Q is the matrix, easily derived from B_F, such that $\underline{F}_0 = B_Q\underline{X}^{\otimes}$. We do know how to find the W that optimizes simple structure in rotated pattern $AB_Q(W^{-1} \otimes W^{-1})$, and that converts directly to a corresponding simple-structured $AB_F W_*^{-1}$. When solution algorithms for (18) or (19) become available, we will pass along this rotation technique as well.

Bottom-line Practicalities.

Unless you are working with data whose latent-source theory has evident distributional implications, you will probably see little reason to give quad-factoring a try until its programming includes the promised routine for identifying factor quad-dependencies. Even so, thinking about what you might do with factor quad-moments may tempt you to take the next step of actually harvesting this information from whatever multivariate data arrays are your current concern. So we had best warn you about a practical limitation on quad-factoring that will likely persist even after QUADFAC's computational procedures have been optimized. This is simply that quad-factoring requires processing of number arrays whose dimensions are roughly proportional to the squares of the corresponding array sizes in 1st-level factoring;

and these quickly become enormous as the number of 1st-level variables becomes appreciable. Not merely does this make for expensive computing, you may well find that the number of variables you wish to quad-factor exceeds the capacity of any mainframe computer to which you have local access. For example, the Univ. of Alberta's Amdahl 5870, with 32 megabytes of memory, will allocate quad-factoring storage space for no more than 15 1st-level variables. The new generation of super-computers should be somewhat more permissive than this, just how much so we are now attempting to ascertain. But even so, the size-window for effective quad-factoring, bounded from below by the number of 1st-level variables required for an informative moment structure and from above by computer capacity, will probably always remain uncomfortably narrow.

To prevail over this window-of-effectiveness bind, applied quad-factoring needs to select its data with exceptional care. For it cannot count on substantial model violations to be averaged out by abundant data redundancies; rather, one or two 1st-level variables that fit poorly may suffice to muddy parameter recovery beyond the limits of useful return. (We do not know this to be so, but see good reason to fear it.) Accordingly, it seems best that empirical quad-factoring research be conducted as a two-stage operation whose first stage is a brutal pruning from one's original battery of data measures those that exhibit conspicuous anomalies --large residuals and method chatter--in preliminary quad-factorings. Specifically, if the maximum number, $\underline{n}_T$, of 1st-level measures to which your computer can allocate quad-factoring storage space is less than the number on which you have sample data, you can scan your full array by fast-QUADFAC runs on assorted $\underline{n}_T$-item subsets thereof. The print-out shows reproduction errors specifically associated with each variable, as well as $\underset{\sim}{u}$-estimates from all four levels of model-(6) utilization described in Appendix B; and this should tell you what pick of at most $\underline{n}_T$ of these items can be passed on to more intensive QUADFAC analysis with minimal manifest model misfit.

And one other admonition: Don't bother to quad-factor small-sample data. Although our studies of QUADFAC performance are still too narrow for authoritative conclusions, we have investigated various levels of sampling noise in arrays of 8 and 12 1st-level variables. (See Appendix D for the 8-variable results.) And whereas source-parameter recovery is near-perfect for artificial data from infinite populations (i.e., no sampling error), and gratifyingly accurate from samples of size 1000, recovery from samples of size 100 is a matter of mirth.

REFERENCES

Anderson, T. W. (1959). Some scaling models and estimation procedures in the latent class model. In Grenander, U. (ed.) Probability and Statistics. New York: Wiley.

Bentler, P. (1983). Some contributions to efficient statistics in structural models; specification and estimation of moment structures. Psychometrika, 48, 493-517.

Etezadi-Amoli, J., & McDonald, R. P. (1983). A second generation nonlinear factor analysis. Psychometrika, 48, 315-342.

Gangestad, S., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. Psychological Review, 92, 317-349.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96, 201-210.

Lazarsfeld, P. F. (1959). Latent structure analysis. In: Koch, S. (ed.), Psychology: A Study of a Science, Vol. 3. New York: McGraw-Hill.

Lazarsfeld, P. F., & Henry, N. W. (1968). Latent Structure Analysis. Boston: Houghton Mifflin.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. The British Journal of Mathematical and Statistical Psychology, 37, 234-251.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. Psychometrika, 27, 397-415.

McDonald, R. P. (1967). Nonlinear factor analysis. Psychometrika Monographs No. 15.

Mooijaart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50, 323-342.

Pollock, D. S. G. (1979). The Algebra of Econometrics. Chichester, England: Wiley & Sons.

Rozeboom, W. W. (1966). Foundations of the Theory of Prediction. Homewood, Ill.: Dorsey Press.

Rozeboom, W. W. (198). HYBALL: A method for subspace-constrained oblique factor rotation. (Forthcoming)

Rozeboom, W. W. (198). Factor indeterminacy: The saga continues. (Forthcoming)

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. Applied Psychological Measurement, 6, 379-396.

Appendix A. Derivation of the quad-error expectations.

Problem: To determine the expected values of $Q(E_0^e)$-elements

$$q_{hi,jk} = \mathcal{E}[t_{hi}e_{jk}] + \mathcal{E}[e_{hi}t_{jk}] + \mathcal{E}[e_{hi}e_{jk}] \quad (A1)$$

wherein

$$e_{ij} = t_i e_j + e_i t_j + e_i e_j \quad .$$

Solution: For all 2nd-level index pairs $\langle hi, jk \rangle$, including index 0 for t_0 and e_0, it follows from (2.2) that the expected product of t_{hi} and e_{jk} has composition

$$\mathcal{E}[t_{hi}e_{jk}] = \mathcal{E}[t_h t_i t_j e_k] + \mathcal{E}[t_h t_i e_j t_k] + \mathcal{E}[t_h t_i e_j e_k] \quad , \quad (A2)$$

while the expected product of e_{hi} and e_{jk} is

$$\begin{aligned}\mathcal{E}[e_{hi}e_{jk}] = {} & \mathcal{E}[t_h e_i t_j e_k] + \mathcal{E}[t_h e_i e_j t_k] + \mathcal{E}[t_h e_i e_j e_k] + \\ & \mathcal{E}[e_h t_i t_j e_k] + \mathcal{E}[e_h t_i e_j t_k] + \mathcal{E}[e_h t_i e_j e_k] + \\ & \mathcal{E}[e_h e_i t_j e_k] + \mathcal{E}[e_h e_i e_j t_k] + \mathcal{E}[e_h e_i e_j e_k] \quad .\end{aligned} \quad (A3)$$

Under the basic error-model's presumption of error independence, together with stipulation of centered scales, most of these terms are zero. But several subcases must be distinguished according to how the various 1st-order subscripts differ therein. The principle of evaluation here is that any term $\mathcal{E}[z_h z_i z_j z_k]$ in (A2,A3) (z either t or e) is zero whenever it contains just one t-component other than t_0 or when any of its e components is either e_0 or occurs just once therein. For example, if $i \neq j$, $\mathcal{E}[t_h e_i e_j t_k] = \mathcal{E}[t_h t_k]\mathcal{E}[e_i]\mathcal{E}[e_j] = 0$ by independence and zero error expectation. And when either $h \neq 0$ or one of i,j,k is distinct from the others, $\mathcal{E}[t_h e_i e_j e_k] = \mathcal{E}[t_h]\mathcal{E}[e_i e_j e_k] = 0$ either because $\mathcal{E}[e_i e_j e_k] = 0$ or, when $i = j = k$ but $h \neq 0$, because centering of Y contrives $\mathcal{E}[t_h] = 0$.

It follows that the only nonzero terms in (A2,A3) for a particular choice of $\langle hi, jk \rangle$ are ones wherein either two e-components each occur twice, one occurs four times, or one occurs three times together with t_0. Accordingly,

$$\mathcal{E}[\underline{t}_0\underline{t}_0\underline{e}_k\underline{e}_k] = \mathcal{E}[\underline{e}_k^2] = \underline{u}_k \ ,$$
$$\mathcal{E}[\underline{t}_i\underline{t}_j\underline{e}_k\underline{e}_k] = \mathcal{E}[\underline{t}_i\underline{t}_j]\mathcal{E}[\underline{e}_k^2] = \underline{c}_{ij}\underline{u}_k \quad \text{if } \underline{i} \neq \underline{j} \ ,$$
$$\mathcal{E}[\underline{t}_i\underline{t}_i\underline{e}_k\underline{e}_k] = \mathcal{E}[\underline{t}_i^2]\mathcal{E}[\underline{e}_k^2] = (\underline{c}_{ii} - \underline{u}_i)\underline{u}_i \ ,$$
$$\mathcal{E}[\underline{e}_i\underline{e}_i\underline{e}_k\underline{e}_k] = \mathcal{E}[\underline{e}_i^2]\mathcal{E}[\underline{e}_k^2] = \underline{u}_i\underline{u}_k \quad \text{if } \underline{i} \neq \underline{k} \ ,$$
$$\mathcal{E}[\underline{t}_0\underline{e}_i\underline{e}_i\underline{e}_i] = \mathcal{E}[\underline{e}_i^3] = \underline{u}_i^{[3]} \ ,$$
$$\mathcal{E}[\underline{e}_i\underline{e}_i\underline{e}_i\underline{e}_i] = \mathcal{E}[\underline{e}_i^4] = \underline{u}_i^{[4]}.$$

Inserting these expectations into (A2,A3) for all the distinctive subcases of 2nd-level index pairs $\langle \underline{hi}, \underline{jk} \rangle$ then yields

$$\mathcal{E}[\underline{t}_{hi}\underline{e}_{jk}] = 0 \quad \text{unless } \underline{j} = \underline{k} \neq 0. \text{ In that case:}$$
$$\mathcal{E}[\underline{t}_{hi}\underline{e}_{jj}] = \mathcal{E}[\underline{t}_h\underline{t}_i]\mathcal{E}[\underline{e}_j^2] = \underline{c}_{hi}\underline{u}_j \quad \text{if } \underline{h} \neq \underline{i} \ ,$$
$$\mathcal{E}[\underline{t}_{ii}\underline{e}_{jj}] = \mathcal{E}[\underline{t}_i^2]\mathcal{E}[\underline{e}_j^2] = (\underline{c}_{ii} - \underline{u}_i)\underline{u}_j \ ,$$

for the elements of $\underset{\sim}{\mathrm{M}}(\underline{T}_0^{\oplus}, \underline{E}_0^{\oplus})$. And the elements of $\underset{\sim}{\mathrm{M}}(\underline{E}_0^{\oplus}, \underline{E}_0^{\oplus})$ are various instantiations o

$$\mathcal{E}[\underline{e}_{0i}\underline{e}_{0j}] = 0 \ ,$$
$$\mathcal{E}[\underline{e}_{hi}\underline{e}_{jk}] = 0 \quad \text{unless either one of } \langle \underline{h}, \underline{i} \rangle \text{ is the same as one of } \langle \underline{j}, \underline{k} \rangle, \text{ or } \underline{h} = \underline{i} \text{ and } \underline{j} = \underline{k}. \text{ In these cases:}$$
$$\mathcal{E}[\underline{e}_{hi}\underline{e}_{hj}] = \mathcal{E}[\underline{t}_i\underline{t}_j]\mathcal{E}[\underline{e}_h^2] = \underline{c}_{ij}\underline{u}_h \quad \text{if } \underline{h}, \underline{i}, \underline{j} \text{ are all distinct} \ ,$$
$$\mathcal{E}[\underline{e}_{hi}\underline{e}_{ii}] = 2\mathcal{E}[\underline{t}_h\underline{t}_i]\mathcal{E}[\underline{e}_i^2] + \mathcal{E}[\underline{t}_h]\mathcal{E}[\underline{e}_i^3] = 2\underline{c}_{hi}\underline{u}_i + \underline{m}_{y_h}\underline{u}_i^{[3]}$$
$$= \left\{ \begin{matrix} 2\underline{c}_{hi}\underline{u}_i & \text{if } 0 < \underline{h} \neq \underline{i} \\ \underline{u}_i^{[3]} & \text{if } 0 = \underline{h} \neq \underline{i} \end{matrix} \right\} (\text{ centered } \underline{Y} \text{ }) \ ,$$
$$\left. \begin{aligned} \mathcal{E}[\underline{e}_{hi}\underline{e}_{hi}] &= \mathcal{E}[\underline{t}_h^2]\mathcal{E}[\underline{e}_i^2] + \mathcal{E}[\underline{e}_h^2]\mathcal{E}[\underline{t}_i^2] + \mathcal{E}[\underline{e}_h^2]\mathcal{E}[\underline{e}_i^2] \\ &= (\underline{c}_{hh} - \underline{u}_h)\underline{u}_i + (\underline{c}_{ii} - \underline{u}_i)\underline{u}_h + \underline{u}_h\underline{u}_i \end{aligned} \right\} \text{if } \underline{h} \neq \underline{i} \ ,$$
$$\mathcal{E}[\underline{e}_{ii}\underline{e}_{jj}] = \mathcal{E}[\underline{e}_i^2]\mathcal{E}[\underline{e}_j^2] = \underline{u}_i\underline{u}_j \quad \text{if } \underline{i} \neq \underline{j} \ ,$$
$$\mathcal{E}[\underline{e}_{ii}\underline{e}_{ii}] = 4\mathcal{E}[\underline{t}_i^2]\mathcal{E}[\underline{e}_i^2] + \mathcal{E}[\underline{e}_i^4] = 4(\underline{c}_{ii} - \underline{u}_i)\underline{u}_i + \underline{u}_i^{[4]} \ .$$

Substitution of these results into (A1) then yields the values of $\underline{q}_{hi,jk}$ reported in (6), p. 9 above.

__Note__.

Even without auxillary assumptions about skew and kurtosis, the quad-factoring error model is appreciably stronger than the "local independence" of errors often postulated by nonlinear item-response theory. (See Anderson, 1959; also McDonald, 1982.) To clarify the difference, define the true-part $\underline{t}_i$ of each data variable $\underline{y}_i$ to be the unrestricted curvilinear regression of $\underline{y}_i$ upon this item-domain's common factors $\underline{F} = \langle \underline{f}_1,\ldots,\underline{f}_r \rangle$, i.e., each subject's value of error variable $\underline{e}_i =_{def} \underline{y}_i - \underline{t}_i$ is his value of $\underline{y}_i$ less the conditional mean of $\underline{y}_i$ among subjects with this same configuration of scores on $\underline{F}$. Then the "local independence" presumption is merely that $\underline{e}_1,\ldots,\underline{e}_n$ (equivalently, $\underline{y}_1,\ldots,\underline{y}_n$) are distributed independently of one another __conditionally__ at each $\underline{F}$-setting; whereas the basic quad-factoring premise is that $\underline{e}_1,\ldots,\underline{e}_n$ are __unconditionally__ independent of each other and of $\underline{T}$, which pretty well requires--not rigorously, but close enough--not merely local independence but also constancy of the conditional distributions of each $\underline{e}_i$ given $\underline{F}$. This is not unreasonable for a data variable $\underline{y}_i$ that is continuous and open-ended; but it cannot strictly hold for any discrete $\underline{y}_i$ (albeit that shouldn't matter much if $\underline{y}$ has decently many scale steps) and may be severely violated if $\underline{y}_i$ has a floor or ceiling approached by appreciably many observations in the data set analyzed.

Even so, none of the expectations $\mathcal{E}[\underline{z}_h\underline{z}_i\underline{z}_j\underline{z}_k]$ ($\underline{z}$ either $\underline{t}$ or $\underline{e}$) developed above under the basic quad-factoring error premise requires full unconditional independencies, and many should be robust under violations of this. We venture that appreciable departures from quad-factoring error model (6) are likely to arise in practice, given a decent approximation to conditional independence, only when floor/ceiling effects are pronounced. In that case, we would anticipate that the terms deviating most from their quad-factoring theoretical values should be the ones of form $\mathcal{E}[\underline{t}_i\underline{e}_i^3]$, $\mathcal{E}[\underline{t}_i^2\underline{e}_i^2]$, and probably $\mathcal{E}[\underline{t}_0\underline{t}_i\underline{e}_i^2]$ if the $\underline{y}_i$-scale is cramped only at one end. If so, the major violations of operational error model (6) should occur in the $\underline{q}_{ii,ii}$ terms, about which all model assumptions are easily waived. Be that as it may, if error-model violations are concentrated in a comparatively small number of error terms $\{\underline{q}_{hi,jk}\}$, these can be picked out by fine-grained assessment of model fit and compensated for by the same solution methodology (Appendix E) that accomodates nonNormal error skew and kurtosis.

Appendix B. QUADFAC programming details.

Starting from an initial estimate $\hat{u}_0^+$ of u^+ (i.e. of u together with none, one, or both of $u^{[4]}$ and $u^{[3]}$ depending on the strength of error-model assumed) and guess r at the number of 1st-level common factors F, QUADFAC iteratively alternates between an improved estimate $\hat{M}_{Ti}$ of true-part quad-moments $M(T_0^*,T_0^*)$ given $\hat{u}_{i-1}^+$ and an improved estimate $\hat{u}_i^+$ of u^+ given $\hat{M}_{Ti}$, generally accompanied by revised estimate $\hat{A}_i$ of factor pattern A and $\hat{M}_{Fi}$ of factor quad-moments $M(F_0^*,F_0^*)$. (Fast QUADFAC does not iterate beyond $i = 1$.) Our main computational tool is classic principal factoring (Eckhard-Young approximation) with certain modifications ensuing from the quad-factor model's special structure. Details follow after a prefatory word about the number of 2nd-level factors.

Quadratic factor dimensionality.

The number $1+r$ of m-complete 1st-level common factors F_0 in (11) is of course one of our major unknowns. But whatever r may be, it fixes the number $1+r^*$ of factors in F_0's bare quadratic development F_0^* as $1+r^* = (r+1)(r+2)/2$, or $r^* = r(r+3)/2$. On first thought, it might seem that r should be the rank of $C(T_0,T_0)$ (equivalently, of $C(T,T)$) identifiable by 1st-level factoring of $C(Y,Y)$, while r^* is the rank of $C(T_0^*,T_0^*)$. But not only does rank-minimizing 1st-level factoring prevailingly underfactor, we have already noted that one benefit of quadratic analysis may well be recovery of factors too weak for detection just in 1st-level data. So we want to encourage solutions of (16/16a) in which r is larger than what would be orthodoxly found by factoring $C(Y,Y)$ with rank-minimizing communalities. And although the number $1+r^*$ of columns in quadratic pattern A_* is rigidly specified by r, the number of appreciably nonzero roots of $C(T_0^*,T_0^*)$ may be considerably less than r^* due to multicollinearities among the 2nd-level factors. This 2nd-level-dependency prospect is not displeasing, for quadratic results are far more interpretively interesting with multicollinearities in F_0^* than without them. In any case, it is important to be clear that the effective rank of $\mathbf{C}(\mathbf{T}_0^*,\mathbf{T}_0^*)$ is just a lower bound on r^*. The only good way to select factor number is to develop solutions over a range of r-choices, including ones larger than what 1st-level factoring would orthodoxly approve, and see how nice is the resultant model fit.

An outline of QUADFAC iterations.

Let $\theta_{EY}(\ )$ be the function defined by equations (6) that maps uniqueness terms $\mathbf{u}^+$ into the corresponding array $Q(\mathbf{E}_0^+)$ of 2nd-level errors that our model expects $\mathbf{u}^+$ to induce in data quad-moments $\mathbf{M}(\mathbf{Y}_0^*,\mathbf{Y}_0^*)$. (The "$\mathbf{Y}$" in this notation serves as reminder that function θ_{EY} includes the 1st-level data covariances as parameters.) That is,

$$Q(\mathbf{E}_0^+) \;=\; \theta_{EY}(\mathbf{u}^+)$$

is error-model (6) writ small. For any fixed r, given an estimate $\hat{\mathbf{u}}_{i-1}^+$ of $\mathbf{u}^+$, we enter the ith cycle of QUADFAC iteration, or more generally a subcycle thereof, by taking $\theta_{EY}(\hat{\mathbf{u}}_{i-1}^+)$ for our running estimate of $\mathbf{Q}(\mathbf{E}_0^+)$, and hence $\mathbf{M}(\mathbf{Y}_0^*,\mathbf{Y}_0^*) - \theta_{EY}(\hat{\mathbf{u}}_{i-1}^+)$ as our corresponding cycle-initiating approximation to true-part quad-moment matrix $\mathbf{M}(\mathbf{T}_0^*,\mathbf{T}_0^*)$. And this cycle/subcycle's yield is a revised estimate $\hat{\mathbf{M}}_{Ti}$ of $\mathbf{M}(\mathbf{T}_0^*,\mathbf{T}_0^*)$ such that the righthand side of

$$\mathbf{M}(\mathbf{Y}_0^*,\mathbf{Y}_0^*) - \theta_{EY}(\hat{\mathbf{u}}_{i-1}^+) \;\simeq\; \hat{\mathbf{M}}_{Ti} \tag{B1}$$

is fitted to its lefthand side under closer proximity to the model's ideal structure than achieved on the left. Solution for $\hat{\mathbf{M}}_{Ti}$ may or may not be accompanied by estimates $\hat{\mathbf{A}}_i$ and $\hat{\mathbf{M}}_{Fi}$ of 1st-level factor pattern $\mathbf{A}$ and factor quad-moments $\mathbf{M}(\mathbf{F}_0^*,\mathbf{F}_0^*)$. When it is, as occurs just at completion of a full cycle, $\hat{\mathbf{A}}_i$, $\hat{\mathbf{M}}_{Fi}$, and a sparce matrix $\mathbf{R}_i$ whose nonzero terms, if any, are corrections of $q_{hi,jk}$-terms whose model-(6) specifications have been suspended, are obtained by fitting the righthand side of

$$\mathbf{M}(\mathbf{Y}_0^*,\mathbf{Y}_0^*) - \theta_{EY}(\hat{\mathbf{u}}_{i-1}^+) \;\simeq\; \hat{\mathbf{A}}_{*i}\hat{\mathbf{M}}_{Fi}\hat{\mathbf{A}}_{*i}' + \mathbf{R}_i \tag{B2}$$

with $\hat{A}_{*i}$ having the $\hat{A}_i$-based structure described by (15), and the triple product on the right giving $\hat{M}_{Ti}$. That is, when (B2) is fitted we put

$$\hat{M}_{Ti} = \hat{A}_{*i}\hat{M}_{Fi}\hat{A}'_{*i} \quad .$$

The nonzero (to-be-fitted) elements of model-relaxation matrix R_i are selected (a) by stipulating one of the three grades of error-model strength, and (b) at control-parameter option, by a subroutine which picks out the <hi,jk>-indices at which previous model fit has most poorly reproduced the data quad-moments. This amounts to waiving the model-(6) constraints on these $q_{hi,jk}$.

Finally, this cycle (or subcycle) derives a new uniqueness estimate u_i by fitting some selection of the component equations in

$$\hat{M}_{Ei} \simeq \theta_{EY}(\hat{u}_i^+) \qquad (B4)$$

where

$$\hat{M}_{Ei} =_{def} M(Y_0^*,Y_0^*) - \hat{M}_{Ti} \quad .$$

This cycle's reproduction of the data quad-moments is then

$$\hat{M}_{Yi} =_{def} \hat{M}_{Ti} + \theta_{EY}(u_i^+) \quad ;$$

and if the fit of approximation $M(Y_0^*,Y_0^*) \simeq \hat{M}_{Yi}$ appreciably improves upon that of the preceding cycle, the iteration continues.

Solution for M_{Ti}. Hypothesizing that the Y-variables have r 1st-level factors entails that the rank of $\hat{M}_{Ti}$ in (B1)-(B3) should not exceed $1+r^*$. An obvious way to achieve this 2nd-level rank constraint is through the Eckhard-Young approximation that replaces by zero all eigenvalues after the $(1+r^*)$th in the eigenstructure decomposition of (B1)'s lefthand side; and with two minor modifications, this is QUADFAC's "coarse" solution of (B1) for $\hat{M}_{Ti}$.

The modifications: (1) We first partial f_0 out of (B1)'s left side before solving the resultant estimate of $C(T_0^*,T_0^*)$ for its first r^* principal axes. (2) M_{Ti} is quad-symmetrized by averaging across elements that quad-symmetry requires to be equal.

This coarse solution for $\hat{M}_{T1}$ does not, however, have explicit decomposition (B3). Ideally, equations (B1)-(B3) should be solved by simultaneously fitting all unknowns on the right in (B2) by some modern structural-modelling algorithm. But pending an effective subroutine for that, QUADFAC's repertoire of "fine" solutions of (B1)-(B3) for $\langle\hat{A}_1,\hat{M}_{F1},R_1\rangle$ and thence M_{T1} proceed as follows: Each variant begins with a solution $\hat{A}_1$ for the 1st-level factor pattern. Fast QUADFAC takes $\hat{A}_1$ to be simply the pattern found by orthodox 1st-level iterated principal factoring of $C(Y,Y)$ expanded to include a row for y_0 and column for f_0.[1] But under the control settings for iterated 2nd-level solutions, each cycle of fine solution for $\hat{A}_1$ first computes a coarse true-quad-moment estimate $\hat{M}_{T1}$ (generally iterated through a small number of coarse subcycles) and solves the estimate of 1st-level true-part covariances $C(T,T)$ embedded therein for the pattern on its first r variance-normalized principal axes. After expansion to include y_0 and f_0, this pattern is then taken for $\hat{A}_1$. However $\hat{A}_1$ is obtained, $\hat{A}_{*1}$ is derived from it by (15), after which $\hat{M}_{F1}$ and R_1 are simultaneously computed to fit (B2) with this fixed $\hat{A}_{*1}$ by the least-squares algorithm described in Appendix E. Since this procedure obtains $\hat{A}_{*1}$ only from the 1st-level part of $\hat{M}_{T1}$, it is clearly suboptimal in principle. Yet it works decently enough with artificial data even when that contains realistic sampling noise; and although our forthcoming structural-modelling alternatives will surely prove superior, the improvement those bring may or may not be appreciable.

<u>Solution for $\hat{u}_1$</u>. There are enormously many ways to solve (B4) for $\hat{u}_1$, but some are far less robust than others. Of the varieties we have tested, the ones that have proved reasonably effective are all classical least-squares fits of over-determined simultaneous linear equations. To examine details, let $\hat{u}_k$ (similarly $\hat{u}_h$) be the kth element of $\hat{u}_1$, i.e., $\hat{u}_k = [\hat{u}_1]_k$. Then from (6), writing unknowns on the left as conventional for simultaneous equations and pretending for tidiness that

[1]The main motivation for fast QUADFAC, namely, bypassing the considerable expense of IMSL's solution for large-matrix eigenstructure, has been largely obviated by the recent release of IMSL:MATHLIB. The new subroutines for eigenstructure therein are faster than before by--incredibly--over an order of magnitude. And they appear more accurate as well.

(B4) is not just an approximation but an identity, each component equation in (B4) that matters for $\hat{\underline{u}}_k$ has the form (up to index permutation) of one of

$$\underline{c}_{hj}\hat{\underline{u}}_k = [\underset{\sim}{M}_{Ei}]_{hj,kk} \qquad (\underline{h},\underline{j},\underline{k} \text{ all distinct; } 1 \le \underline{h} < \underline{j}) \qquad (B5.1)$$

$$3\underline{c}_{hk}\hat{\underline{u}}_k = [\underset{\sim}{M}_{Ei}]_{hk,kk} \qquad (1 \le \underline{h} < \underline{k}) \qquad (B5.2)$$

$$\hat{\underline{u}}_k = [\underset{\sim}{M}_{Ei}]_{00,kk} \qquad (B5.3)$$

$$\underline{c}_{hh}\hat{\underline{u}}_k + \underline{c}_{kk}\hat{\underline{u}}_h - \hat{\underline{u}}_h\hat{\underline{u}}_k = [\underset{\sim}{M}_{Ei}]_{hh,kk} \qquad (1 \le \underline{h} < \underline{k}) \qquad (B5.4)$$

$$6\underline{c}_{kk}\hat{\underline{u}}_k - 3\hat{\underline{u}}_k^2 = [\underset{\sim}{M}_{Ei}]_{kk,kk} \qquad (\text{if error kurtosis is Normal}) \qquad (B5.5)$$

where standard scaling in QUADFAC practice puts $\underline{c}_{hh} = \underline{c}_{kk} = 1$ in (B5.4,5). All of these except (B5.4) and (B5.5) are linear in their unknowns; and that becomes true of the latter as well if we replace $\hat{\underline{u}}_h\hat{\underline{u}}_k$ and $3\hat{\underline{u}}_k^2$ therein by their approximations computed from our last estimate of $\underset{\sim}{u}$ (i.e. either $\hat{\underset{\sim}{u}}_{i-1}$ or the most recent estimate reached by iterating (B5)'s linear-equations solution). Because the full array of equations (B5) vastly overdetermines $\hat{\underset{\sim}{u}}_i$, it is feasible to solve only selected subarrays in hope of avoiding quad-moments particularly susceptable to poor fit. At present, QUADFAC provides alternative solutions for $\hat{\underset{\sim}{u}}_i$ from four nested selections from (B5). In order of increasing inclusion, these are:

<u>Selection 1</u>. Just the equations of form (B5.3). This is a traditional 1st-level uniqueness solution, and the one used by fast QUADFAC.

<u>Selection 2</u>. All the equations of form (B5.1,2,3). This subarray has a direct least-squares solution for each $\hat{\underline{u}}_k$ separately.

<u>Selection 3</u>. All of equations (B5) except those of form (B5.5). This subset ignores the kurtosis estimates in $\underset{\sim}{M}_{Ei}$, which are usually much larger than other terms in $\underset{\sim}{M}_{Ei}$ and suffer the greatest sampling variance.

<u>Selection 4</u>. All equations (B5), including subarray (B5.5). This is appropriate only when Normal error kurtosis is presumed.

Our artificial-data studies of QUADFAC performance (see Appendix D) have not yet discerned any clear superiority order on these options, albeit Selection 4 is clearly inadvisable for data suspected to be appreciably contaminated by floor/ceiling effects. Although any one QUADFAC run iterates just one of these solution options, it prints out the **u**-estimates from all four Selections on each iteration cycle.

Appendix C. Fragments of the theory of quadratic spaces.

In a broad sense, the quadratic functions of variables $\underline{X} = \langle \underline{x}_1,\ldots,\underline{x}_n \rangle$ are all those of form $\phi(\underline{X}) = \underline{a}_0 + \sum_{i=1}^{n} \underline{a}_i\underline{x}_i + \sum_{j=1}^{n}\sum_{k=j}^{n} \underline{a}_{jk}\underline{x}_j\underline{x}_k$. But here we shall adopt the narrower usage wherein the quadratic functions of variables $\underline{X}$ are just the ones of homogeneous form $\phi(\underline{X}) = \sum_{j=1}^{n}\sum_{k=j}^{n} \underline{a}_{jk}\underline{x}_j\underline{x}_k$. (As will be noted, the broad sense is recoverable as a special case under the narrow one.) By the linear space, $\mathcal{L}_X$, spanned by a tuple $\underline{X} = \langle \underline{x}_1,\ldots,\underline{x}_n \rangle$ of variables we shall mean, as usual, the set of all homogeneous linear combinations of the $\underline{X}$-variables, i.e., all functions of form $\phi(\underline{X}) = \sum_{i=1}^{n} \underline{a}_i\underline{x}_i$. Let us say that tuple $\underline{X}$ of variables is (**implicitly**) **complete** iff the unit variable is in $\mathcal{L}_X$, and that $\underline{X}$ is m(anifestly)-complete iff the unit variable is one of those in tuple $\underline{X}$. Whenever we write $\underline{X} = \langle \underline{x}_0, \underline{x}_1,\ldots,\underline{x}_n \rangle$ for a tuple of non-error variables, i.e., with the tuple's indexing starting with 0 rather than 1, we presume $\underline{X}$ to be m-complete with $\underline{x}_0$ the unit variable. (Error tuples $\underline{E}_0$, $\underline{E}_0^+$, and $\underline{E}_0^{\theta}$ remain exceptions to this rule, but will not be mentioned in this Appendix.) The space $\mathcal{L}_{X_0}$ linearly spanned by the m-completion $\underline{X}_0 = \langle \underline{x}_0, \underline{x}_1,\ldots,\underline{x}_n \rangle$ of $\underline{X}$ comprises all linear combinations of $\langle \underline{x}_1,\ldots,\underline{x}_n \rangle$ that include additive constants. And since $\underline{x}_0\underline{x}_i = \underline{x}_i$ ($\underline{i} = 0,1,\ldots,\underline{n}$), the quadratic functions of $\underline{X}_0$ in the narrow (homogeneous) sense include all quadratic functions of $\underline{X}$ in the broad sense that admits linear terms and additive constants. Hence in particular, $\mathcal{L}_{X_0} \subset \mathcal{Q}_{X_0}$.

It is often insightful to express quadratic functions $\phi(\underline{x}_1,\ldots,\underline{x}_n) = \sum_{i=1}^{n}\sum_{j=i}^{n} \underline{a}_{ij}\underline{x}_i\underline{x}_j$ in matrix form $\phi(\underline{X}) = \underline{X}'\underset{\sim}{Q}\underline{X}$, wherein $\underline{X} = \langle \underline{x}_1,\ldots,\underline{x}_n \rangle$ is algebraically a column vector and $\underset{\sim}{Q}$ is the $\underline{n} \times \underline{n}$ symmetric matrix whose $\underline{ij}$th element is $\underline{a}_{ii}$ if $\underline{i} = \underline{j}$, $\underline{a}_{ij}/2$ if $\underline{i} < \underline{j}$, and $\underline{a}_{ji}/2$ if $\underline{i} > \underline{j}$. Then the quadratic space, $\mathcal{Q}_X$, generated by variables $\underline{X}$ is the set of all functions $\{\phi_Q(\underline{X}) = \underline{X}'\underset{\sim}{Q}\underline{X}$: $\underset{\sim}{Q}$ any $\underline{n} \times \underline{n}$ symmetric real matrix$\}$. $\mathcal{Q}_X$ is also a space in the standard linear sense, since all homogeneous linear combinations of functions in $\mathcal{Q}_X$ are themselves in $\mathcal{Q}_X$. Indeed, $\mathcal{Q}_X$ is the space $\mathcal{L}_{X*}$ linearly spanned by the bare quadratic development $\underline{X}^*$ of $\underline{X}$, and is hence linearly spanned also by X^*. And if X is a basis for its linear space $\mathcal{L}_X$, X^* fails to be a basis for $\mathcal{Q}_X$ just in case, for some tuple Z of variables in $\mathcal{L}_X$, all Z-points lie on a hyperbolic surface.

Proof. Variables X^* contain a homogeneous linear dependency (relative to a given population in which X is distributed) iff $X'QX = 0$ for some nonzero symmetric Q. By virtue of its symmetry, Q can always be decomposed as $Q = T'DT$ where T is orthonormal and D is diagonal though perhaps not positive definite. Hence if $Z =_{def} TX$, $X'QX = 0$ iff $Z'DZ = 0$, i.e. iff $\sum_i d_i z_i^2 = 0$ for the n roots (diagonal elements) $\{d_i\}$ of D. If all nonzero roots of D, say $d_1, \ldots, d_r$ $(r \leq n)$, have the same sign, it follows for each $i = 1, \ldots, r$ that $z_i^2 = 0$ and hence $z_i = 0$—which is to say that linear X-space is at most $(n-r)$-dimensional contrary to assumption that X is a basis for $\mathcal{L}_X$. Alternatively, if some of the $r \leq n$ nonzero D-roots are opposed in sign, $Z'DZ = 0$ is the equation for a hyperbolic surface in the subspace of $\mathcal{Q}_X$ spanned by the first r variables in Z. And the bi-directionality of this argument is plain. □

Finally, it is of fundamental importance for quadratic factoring that if X and Z linearly span the same space $\mathcal{L}_X = \mathcal{L}_Z$, then, regardless of any linear dependencies in X or Z, X^* and Z^* both span the same quadratic space $\mathcal{Q}_X = \mathcal{L}_{X*} = \mathcal{L}_{Z*} = \mathcal{Q}_Z$.

Proof. Suppose that Z and X span the same linear space even though the number m of variables in Z may differ from the number n in X. Then there exist not-necessarily-unique coefficient matrices A and B of order $m \times n$ and $n \times m$, respectively, such that $Z = AX$ and $X = BZ$. So if Q_m and Q_n are respectively any $m \times m$ and $n \times n$ symmetric quadratic-coefficient matrices, $Z'Q_mZ = X'(A'Q_mA)X$ and $X'Q_nX = Z'(B'Q_nB)Z$.

It is not generally the case, however, that if variables X are orthogonal to variables Z, then $\mathcal{Q}_X$ is orthogonal to $\mathcal{Q}_Z$. (In this paper, we understand "orthogonality" in its generic sense of zero 2nd-order moments or zero vector products, not in its special sense of zero 1st-level covariances.) In particular, if means have been partialled out of $\underline{X}$, i.e. if $\underline{X}$ is orthogonal to $\mathbf{x}_0$, most variables in $\mathcal{Q}_X$ still retain nonzero means. (Recall that any variable's mean equals its mean product with the unit variable.) For this reason, quad-factoring cannot partial out 1st-level means and thereafter work exclusively with covariances as does traditional 1st-level factoring.

The rudiments of quadratic-function theory needed for present purposes can be expressed with powerful elegance in the language of tensor algebra. Central to this is the Kronecker product, $\mathbf{B} \otimes \mathbf{A}$, of any two matrices $\mathbf{A}$ and $\mathbf{B}$. If $\mathbf{B}$ is $m \times n$ and $\mathbf{A}$ is $r \times s$, $\mathbf{B} \otimes \mathbf{A}$ is defined to be the $mr \times ns$ matrix so partitionable in correspondence with the elements $\{b_{ij}\}$ of $\mathbf{B}$ that for each $i = 1,\ldots,m$ and $j = 1,\ldots,n$, the ijth block (i.e. submatrix) in $\mathbf{B} \otimes \mathbf{A}$ is $b_{ij}\mathbf{A}$. We also need the vec operator that transforms any matrix $\mathbf{A}$ into a super-column of $\mathbf{A}$'s columns. Specifically, when $r \times s$ matrix $\mathbf{A}$ is partitioned by columns as $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_s]$, $\underline{\text{vec}}(\mathbf{A})$ is the order-rs column vector

$$\underline{\text{vec}}([\mathbf{a}_1\ \ldots\ \mathbf{a}_s]) = \begin{bmatrix} \mathbf{a}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_s \end{bmatrix}$$

For inclusion of this operator in formulas, however, we prefer Pollock's (1979, p. 68) more compact notation

$$\mathbf{A}^c =_{\text{def}} \underline{\text{vec}}(\mathbf{A}) ,$$

wherein the superscript is an obvious heurism for "column."

Some basic consequences of these definitions that hold whenever the matrices at issue conform are

(i) $$(\mathbf{AXB}')^c = (\mathbf{B} \otimes \mathbf{A})\mathbf{X}^c .$$

(ii) $$(\mathbf{ab}')^c = \mathbf{b} \otimes \mathbf{a} \quad (\ \mathbf{a} \text{ and } \mathbf{b} \text{ any column vectors}\ ) .$$

(iii) $$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C} , \quad \mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C} .$$

(iv) $$(\mathbf{B} \otimes \mathbf{A})' = \mathbf{B}' \otimes \mathbf{A}' .$$

(v) $$(\mathbf{B} \otimes \mathbf{A})(\mathbf{D} \otimes \mathbf{C}) = \mathbf{BD} \otimes \mathbf{AC} .$$

(vi) $\quad I_{(m)} \circledast I_{(n)} = I_{(mn)}$ $\quad$ ($I_{(k)}$ the $k \times k$ Identity matrix) .

(vii) $\quad (B \circledast A)^L = B^L \circledast A^L$ $\quad$ (left-invertible A and B) .

A matrix A is left-invertible iff its rank equals its column-order, in which case A has a not-necessarily-unique left-inverse $A^L = (A'A)^{-1}A'$ by which $A^LA = I$. The condition for left-inverting $A \circledast B$ is immediate from (v,vi).

(viii) If $A_1 \circledast A_2 = B_1 \circledast B_2$ with A_1, A_2 of the same order respectively as B_1, B_2, then $A_1 = \iota B_1$ and $A_2 = \iota B_2$ where either $\iota = 1$ or $\iota = -1$.

Hence in particular, $A \circledast A = B \circledast B$ iff either $A = B$ or $A = -B$.

Continuing to treat variables $X = \langle x_1, \ldots, x_n \rangle$ as a column vector, we can now write the full quadratic development $X^{\circledast}$ of X as the order-n^2 column vector of pairwise product-variables

$$X^{\circledast} =_{def} (XX')^c = X \circledast X \quad .$$

Each variable x_{ij} ($i,j = 1,\ldots,n$) in tuple $X^{\circledast}$ has composition $x_{ij} = x_i x_j$ and is also one of the 2nd-level variables in array $X^* = \{x_i x_j : i,j = 1,\ldots,n;\ i \le j\}$. The only difference between $X^{\circledast}$ and X^* is that each x_{ij} occurs twice in $X^{\circledast}$ (with permuted subscript) if $i \neq j$. Observe that any quadratic composites $\{g_k =_{def} X'Q_kX\}$ of variables X can be organized as

$$g_k = X'Q_kX = (X'Q_kX)^c = (X' \circledast X')Q_k^c = Q_k^{c\prime}(X \circledast X) = Q_k^{c\prime}X^{\circledast} \ ,$$

and collected into a column vector $G = \langle g_1, g_2, \ldots \rangle$ of variables having classic linear multivariate form

$$G = W_Q X^{\circledast} \qquad ([W_G]_{k.} =_{def} Q_k^{c\prime},\ k = 1,2,\ldots) \ .$$

As a special case of this format, for any tuple of variables $Z = AX$ in the linear space of X, the full quadratic development of Z is linearly determined by that of X according to

$$Z^{\circledast} \;=\; (Z \circledast Z) \;=\; AX \circledast AX \;=\; (A \circledast A)(X \circledast X) \;=\; (A \circledast A)X^{\circledast} \qquad (Z = AX) .$$

And if A is of full column-rank and so has a left-inverse, this dependency of $Z^{\circledast}$ upon $X^{\circledast}$ can be inverted as

$$X^{\circledast} \;=\; (A \circledast A)^{L} Z^{\circledast} \;=\; (A^{L} \circledast A^{L}) Z^{\circledast} \qquad (Z = AX,\; A^{L} = (A'A)^{-1}A')$$

to reclaim $X^{\circledast}$ from $Z^{\circledast}$. A necessary condition for A^{L} to exist is for Z to span $\mathcal{L}_X$; and that together with X's being a basis for $\mathcal{L}_X$ is also sufficient. Unhappily, the situation is messier if X is not a basis for $\mathcal{L}_X$; for then there are many coefficient matrices $\{A_i\}$ such that $Z = A_i X$, and not all of these have left-inverses even when Z spans $\mathcal{L}_X$. But some do--which is to say that so long as Z spans $\mathcal{L}_X$, there always exists at least one coefficient matrix A such that $Z = AX$ and $X = A^{L}Z$; whence also $Z^{\circledast} = (A \circledast A)X^{\circledast}$ is invertible as $X^{\circledast} = (A^{L} \circledast A^{L})Z^{\circledast}$. (See Rozeboom, , for proof of this and other cheerful facts about left-invertible factor patterns henceforth taken for granted here.)

Not merely do these formulas concisely describe how linear relations among 1st-level variables unfold into linear relations among quadratic functions thereof, they also show in principle how to analyze linear dependencies in a quadratic space into relations among axes in the underlying linear space. Let $Z = \langle z_1,\ldots,z_n \rangle$ be an n-tuple of variables (which may or may not be m-complete) whose 4th-order moments we have identified either by direct computation when Z comprises empirical measures or, when the Z-variables are true-parts, by correction for 2nd-level error as described elsewhere. And suppose that study of these moments has revealed that $M(Z^{\circledast}, Z^{\circledast})$, i.e. $\mathcal{E}[Z^{\circledast} Z^{\circledast\prime}]$, has a decomposition of quadratic form

$$M(Z^{\circledast}, Z^{\circledast}) \;=\; (A \circledast A)\, M_b\, (A \circledast A)'$$

for some $n \times r$ matrix A $(r \le n)$ and some $r^2 \times r^2$ matrix M_b. If A is of full column-rank r, so that A^{L} and hence $(A \circledast A)^{L}$ exist, there is just one tuple of variables G such that $Z^{\circledast} = (A \circledast A)G$, namely $G = (A \circledast A)^{L} Z^{\circledast}$, whence also $M(G,G) = M_g$. Moreover, these $\underline{r}^2$ 2nd-level $\underline{G}$-factors of $\underline{Z}^{\circledast}$ are immediately identifiable as the quadratic development of 1st-level factor $\underline{r}$-tuple

$$\underline{F} =_{def} \underset{\sim}{A}^L\underline{Z} ,$$

since

$$\underline{G} = (\underset{\sim}{A}\circledast\underset{\sim}{A})^L\underline{Z}^{\circledast} = (\underset{\sim}{A}^L\circledast\underset{\sim}{A}^L)(\underline{Z}\circledast\underline{Z}) = \underset{\sim}{A}^L\underline{Z}\circledast\underset{\sim}{A}^L\underline{Z} = \underline{F}\circledast\underline{F} = \underline{F}^{\circledast} .$$

Insomuch as $\underline{F}^{\circledast} = \underline{G}$, the $\underline{F}$ so identified has 4th-order moments $\underset{\sim}{M}(\underline{F}^{\circledast},\underline{F}^{\circledast}) = \underset{\sim}{M}(\underline{G},\underline{G}) = \underset{\sim}{M}_b$ and reproduces the $\underline{Z}$-information as

$$\underline{Z} = \underset{\sim}{A}\underline{F} , \quad \underline{Z}^{\circledast} = (\underset{\sim}{A}\circledast\underset{\sim}{A})\underline{F}^{\circledast} , \quad \underset{\sim}{M}(\underline{Z}^{\circledast},\underline{Z}^{\circledast}) = (\underset{\sim}{A}\circledast\underset{\sim}{A})\underset{\sim}{M}(\underline{F}^{\circledast},\underline{F}^{\circledast})(\underset{\sim}{A}\circledast\underset{\sim}{A})' ,$$

just as wanted of a simultaneous factor solution at both levels. Finally, note that except for reflection, this $\underline{F}$ is the only $\underline{r}$-tuple of $\underline{Z}$'s 1st-level factors whose quadratic development so reproduces $\underset{\sim}{M}(\underline{Z}^{\circledast},\underline{Z}^{\circledast})$ from A. Specifically, if $\underline{F}_a$ is any basis for linear $\underline{Z}$-space that factors $\underline{Z}$ as $\underline{Z} = \underset{\sim}{B}_a\underline{F}_a$ for some 1st-level pattern $\underset{\sim}{B}_a$ while also $\underline{Z}^{\circledast} = (\underset{\sim}{A}\circledast\underset{\sim}{A})\underline{F}_a^{\circledast}$, then $\underset{\sim}{B}_a$ differs from $\underset{\sim}{A}$ by at most a reflection of some of its columns. Indeed, only for a bizarre distribution can it fail that either $\underset{\sim}{B}_a = \underset{\sim}{A}$ and hence $\underline{F}_a = \underset{\sim}{A}^L\underline{Z} = \underline{F}$, or $\underset{\sim}{B}_a = -\underset{\sim}{A}$ and hence $\underline{F}_a = (-\underset{\sim}{A}^L)\underline{Z} = -\underline{F}$.

Proof. Premises $\underline{Z} = B_a\underline{F}_a$ and $\underline{Z}^{\circledast} = (A\circledast A)\underline{F}_a^{\circledast}$ have the immediate consequence

$$(A\circledast A)\underline{F}_a^{\circledast} = (B_a\circledast B_a)\underline{F}_a^{\circledast} \quad . \tag{C1}$$

Were $\underline{F}_a^{\circledast}$ a basis for $\mathcal{Q}_Z$ it would follow from (C1) that $\underline{A}\circledast\underline{A} = \underline{B}_a\circledast\underline{B}_a$, whence the theorem would be immediate under principle (viii); however, we have already explained why not even $\underline{F}_a^*$, much less $\underline{F}_a^{\circledast}$, is generally a basis for $\mathcal{Q}_Z$ despite $\underline{F}_a$'s being one for $\mathcal{L}_Z$. Nevertheless, if $\underset{\sim}{f}$ is any column-vector of scores on $\underline{F}_a$ for some member of the population $\underline{P}$ in which the distribution of $\underline{Z}$ is at issue, it follows from (C1) under (v) that $\underset{\sim}{A}\underset{\sim}{f}\circledast\underset{\sim}{A}\underset{\sim}{f} = \underset{\sim}{B}_a\underset{\sim}{f}\circledast\underset{\sim}{B}_a\underset{\sim}{f}$ and hence, under (viii), that

$$\underset{\sim}{A}\underset{\sim}{f} = \iota\underset{\sim}{B}_a\underset{\sim}{f} \qquad (\ \iota = 1 \text{ or } \iota = -1\ ) \tag{C2}$$

Now, $\underline{F}_a$ is by stipulation a basis for $\mathcal{L}_Z$, insuring the existence both of $\underset{\sim}{A}^L$ and of $\underline{r}$ linearly independent score-tuples on $\underline{F}_a$ in $\underline{P}$. So there must also exist an $\underline{r} \times \underline{r}$ nonsingular matrix $\mathbf{S}$ whose columns are score-tuples on $\underline{F}_a$ in $\underline{P}$ and, in light of (C2), a diagonal matrix $\mathbf{D}_u$ each root of which is either 1 or -1--call any such $\mathbf{D}_u$ a "reflection" matrix--such that

$$\mathbf{AS} = \mathbf{B}_a\mathbf{S}\mathbf{D}_u = \mathbf{B}_a\mathbf{A}^L\mathbf{A}\mathbf{S}\mathbf{D}_u \quad .$$

Hence, since $\mathbf{D}_u^2 = \mathbf{I}$,

$$\mathbf{AS} = \mathbf{B}_a\mathbf{A}^L(\mathbf{B}_a\mathbf{S}\mathbf{D}_u)\mathbf{D}_u = \mathbf{B}_a\mathbf{A}^L\mathbf{B}_a\mathbf{S} \quad ,$$

which postmultiplication by $\mathbf{S}^{-1}$ reduces to

$$\mathbf{A} = \mathbf{B}_a\mathbf{A}^L\mathbf{B}_a \quad . \tag{C3}$$

And premultiplication of (C3) by $\mathbf{A}^L$ shows that $(\mathbf{A}^L\mathbf{B}_a)^2 = \mathbf{I}$ or, equivalently,

$$\mathbf{A}^L\mathbf{B}_a = \mathbf{D}_v \tag{C4}$$

for some reflection matrix $\mathbf{D}_v$. Finally, insertion of (C4) first into (C3) and then into the premultiplication of (C1) by $\mathbf{A}^L \circledast \mathbf{A}^L$ yields

$$\mathbf{A} = \mathbf{B}_a\mathbf{D}_v \tag{C5}$$

and

$$\underline{F}_a \circledast \underline{F}_a = (\mathbf{A}^L\mathbf{B}_a \circledast \mathbf{A}^L\mathbf{B}_a)\underline{F}^{\circledast} = (\mathbf{D}_v \circledast \mathbf{D}_v)(\underline{F}_a \circledast \underline{F}_a) = \mathbf{D}_v\underline{F}_a \circledast \mathbf{D}_v\underline{F}_a$$

or, equivalently,

$$\underline{F}_a\underline{F}_a' = (\mathbf{D}_v\underline{F}_a)(\mathbf{D}_v\underline{F}_a)' \quad . \tag{C6}$$

If all roots of $\mathbf{D}_v$ have the same sign, then either $\mathbf{D}_v = \mathbf{I}$ or $\mathbf{D}_v = -\mathbf{I}$, whence by (C5) either $\mathbf{B}_a = \mathbf{A}$ or $\mathbf{B}_a = -\mathbf{A}$. Otherwise, $\underline{F}_a$ partitions into two non-null subarrays $\underline{F}_1$ and $\underline{F}_2$ such that, from (C6), $\underline{F}_1\underline{F}_2' = -\underline{F}_1\underline{F}_2'$. This occurs just under the bizarre distributional circumstance that every tuple of scores on $\underline{F}_a$ occurrent in $\underline{P}$ is all zero either on subarray $\underline{F}_1$ or on subarray $\underline{F}_2$. □

The essential point to be taken from this is that so long as we do not stray from left-invertible factor patterns, there is only one modest obstacle to achieving alignment between 1st-level and 2nd-level factor solutions. Quad-factoring's alignment problem is this: When we set out to interpret some decomposition $\mathbf{M}(\underline{Z}^{\circledast},\underline{Z}^{\circledast})$

$= BM_bB'$ of the 4th-order Z-moments, we know that if B is left-invertable then there exist variables G in $\mathcal{Q}_Z$ such that $Z^{\circledast} = BG$ and $M(G,G) = M_b$. But we also know that these G-variables are in turn quadratic functions of whatever 1st-level factor array F we may choose as axes for linear Z-space. Insomuch as the 1st-level Z-moments also have a factoring $M(Z,Z) = AM_aA'$ for any such F, with $Z = AF$ and $M(F,F) = M_a$, how can we extract some F, A, and the specific quadratic determination of G by F from our 2nd-level analysis and reconcile these with whatever might emerge just from the 1st-level analysis of $M(Z,Z)$? Although we have no operational answer to this question for an arbitrary 2nd-level factor pattern, all falls nicely into place if we can only manage to structure the pattern matrix in $M(Z^{\circledast},Z^{\circledast}) = BM_bB'$ as $B = A \circledast A$ for some left-invertable A. For then, as just shown, $F =_{def} A^LZ$ is a 1st-level factor solution that also analyzes the 2nd-level factors in $Z^{\circledast} = BG = (A \circledast A)G$ as $G = F^{\circledast}$, and the G-moments $M(G,G) = M_b$ as the 4th-order moments $M(F^{\circledast},F^{\circledast}) = M(G,G)$ of F. In theory, this F can then be rotated into any 1st-level factor solution we might develop just from $M(Z,Z)$; in practice, failure of such rotations to achieve perfect matches tells us something about differences in what can be recovered from noisy data by 1st-level vs. 2nd-level factoring.

When Z is m-complete, notably when in practice Z is true-part $(n+1)$-tuple $T_0 = \langle t_0, T\rangle$, we have no need for separate factor solutions on both levels insomuch as the 2nd-level analysis embeds a 1st-level one. But there is still an alignment problem in this case. For when 2nd-level true-moment decomposition $M(T_0^{\circledast},T_0^{\circledast}) = BM_bB'$ reveals factors G $(= B^LT_0^{\circledast})$ in $\mathcal{Q}_{T_0}$ such that $T_0^{\circledast} = BG$, even though the first $n+1$ variables in $T_0^{\circledast}$ are 1st-level array T_0, the G-factors to which the first $n+1$ rows of B give nonzero weight are not necessarily in $\mathcal{L}_{T_0}$—especially not if B is developed by something like orthodox principal factoring. Nevertheless, if we require B to have structure $B = A \circledast A$ with $\langle 1,0,\ldots,0\rangle$ for A's 1st row, we insure that $G = F_0 \circledast F_0$ for some $(r+1)$-tuple F_0 of axes in $\mathcal{L}_{T_0}$ commencing with the unit variable. And the leading $(n+1) \times (n+1)$ submatrix in B $(= A \circledast A)$ is then also the 1st-level pattern of T_0 on F_0.

Appendix E. Least-squares Solution for Special Terms in Factor-moment Estimation.

In structural modelling, when we conjecture that tuples $\underline{Y}_a$ and $\underline{Y}_b$ of manifest variables are structurally dependent on source variables $\underline{F}_a$ and $\underline{F}_b$, respectively, according to structural equations

$$\underline{Y}_a = \mathbf{A}\underline{F}_a + \underline{E}_a , \qquad \underline{Y}_b = \mathbf{B}\underline{F}_b + \underline{E}_b ,$$

wherein $\langle \underline{E}_a, \underline{E}_b \rangle$ are residuals, need sometimes arises to estimate $\mathbf{M}(\underline{F}_a, \underline{F}_b)$ given prior estimates of $\langle \mathbf{A}, \mathbf{B} \rangle$ and a more-or-less complex structure on the otherwise unknown contribution to $\mathbf{M}(\underline{Y}_a, \underline{Y}_b)$ of $\langle \underline{E}_a, \underline{E}_b \rangle$. (In QUADFAC applications, $\underline{Y}_a = \underline{Y}_b = \underline{Y}_0^*$, $\mathbf{A} = \mathbf{B} = \mathbf{A}_*$, $\underline{F}_a = \underline{F}_b = \underline{F}_0^*$, and $\underline{E}_a = \underline{E}_b = \underline{E}_0^+$.) To keep notation simple, let $\mathbf{M}_0$ be manifest-moment matrix $\mathbf{M}(\underline{Y}_a, \underline{Y}_b)$ while $\mathbf{M}_F$ is factor-moment matrix $\mathbf{M}(\underline{F}_a, \underline{F}_b)$. Then our model for $\mathbf{M}_0$ is

$$\mathbf{M}_0 = \mathbf{A}\mathbf{M}_F\mathbf{B}' + \mathbf{Q}_0 ,$$

where $\mathbf{Q}_0$ $(= \mathbf{A}\mathbf{M}(\underline{F}_a, \underline{E}_b) + \mathbf{M}(\underline{E}_a, \underline{F}_b)\mathbf{B}' + \mathbf{M}(\underline{E}_a, \underline{E}_b))$ is a matrix of residuals. (In QUADFAC applications, $\mathbf{Q}_0 = \mathbf{Q}_0^+$.)

Suppose that when we seek to extract $\mathbf{M}_F$ from $\mathbf{M}_0$, pattern matrices $\mathbf{A}$ and $\mathbf{B}$ have already been estimated while residual matrix $\mathbf{Q}_0$ is analyzable as $\mathbf{Q}_0 = \mathbf{Q}_1 + \mathbf{Q}$ where $\mathbf{Q}_1$ is numerically fixed and $\mathbf{Q}$ is a sparce matrix whose nonzero elements are open parameters. (In QUADFAC applications, $\mathbf{Q}_0$ is specified by the strong version of error model (6) from the latest estimate of 1st-level uniquenesses $\mathbf{u}$, while $\mathbf{Q}$ contains to-be-estimated correction terms at quadratic-index positions $\{\langle \underline{ii}, \underline{ii} \rangle ; \underline{i} = 1,\ldots,\underline{n}\}$ for waiving presumption of Normal error kurtosis, as well as at quadratic-index positions $\{\langle 0\underline{i}, \underline{ii} \rangle ; \underline{i} = 1,\ldots,\underline{n}\}$ if zero error skew is to be waived.) Our task is to find $\mathbf{M}_F$ and the nonzero elements of $\mathbf{Q}$ that optimize the fit of

$$\mathbf{M}_0 \simeq \mathbf{A}\mathbf{M}_F\mathbf{B}' + (\mathbf{Q}_1 + \mathbf{Q}) . \qquad \text{(E1)}$$

Although this problem can be routinely solved by modern structural-modelling methods, it also has an explicit least-squares solution when $\mathbf{A}$ and $\mathbf{B}$ have left-inverses as follows: Let $\sigma = \{\underline{hi}\}$ be the set of index-pairs that pick out the nonzero elements of $\mathbf{Q}$, i.e., $[\mathbf{Q}]_{hi}$ is a free parameter in $\mathbf{Q}$ just in case $\underline{hi}$ is in set σ. Also, write

$$\mathbf{M}_1 =_{def} \mathbf{M}_0 - \mathbf{Q}_0 \ ,$$

$$\mathbf{A}^L =_{def} (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \ , \qquad \mathbf{P}_A =_{def} \mathbf{A}\mathbf{A}^L \ ,$$

$$\mathbf{B}^L =_{def} (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' \ , \qquad \mathbf{P}_B =_{def} \mathbf{B}\mathbf{B}^L \ .$$

Then putting $\mathbf{E} = \mathbf{M}_0 - (\mathbf{A}\mathbf{M}_F\mathbf{B}' + \mathbf{Q}_1 + \mathbf{Q}) = \mathbf{M}_1 - (\mathbf{A}\mathbf{M}_F\mathbf{B}' + \mathbf{Q})$ for the matrix of approximation errors in (E1), differentiating traditional loss-function $Tr[\mathbf{E}\mathbf{E}']$ $\underline{wrt}$ the unknowns in $\mathbf{M}_F$ and $\mathbf{Q}$, and solving for its minimum shows that the least-squares optimization of (E1) is the solution for $\langle \mathbf{M}_F, \mathbf{Q} \rangle$ in simultaneous equations

$$\mathbf{M}_F = \mathbf{A}^L(\mathbf{M}_1 - \mathbf{Q})\mathbf{B}^{L'} \qquad \text{(E2)}$$

$$[\mathbf{Q} - \mathbf{P}_A\mathbf{Q}\mathbf{P}_B']_{hi} = [\mathbf{M}_1 - \mathbf{P}_A\mathbf{M}_1\mathbf{P}_B']_{hi} \qquad (\underline{hi} \in \sigma) \ . \qquad \text{(E3)}$$

(It seems conceptually helpful to leave the transpose marker on $\mathbf{P}_B$ here even though $\mathbf{P}_A$ and $\mathbf{P}_B$ are symmetric. Proof of this solution is available on request.) (E3) comprises a set of simultaneous linear equations just for the σ-indexed unknowns in $\mathbf{Q}$ without involvement of $\mathbf{M}_F$; and once $\mathbf{Q}$ is found from (E3), its insertion into (E2) yields an explicit solution for $\mathbf{M}_F$.

To solve (E3), let $\mathbf{q}$ be the column vector of the unknown $\mathbf{Q}$-elements arbitrarily ordered as $\langle \ldots, (\underline{hi}), \ldots \rangle$, where $(\underline{hi})$ is the single-index position in $\mathbf{q}$ of doubly indexed $\mathbf{Q}$-element $[\mathbf{Q}]_{hi}$. For each of these $\mathbf{q}$-indices $(\underline{hi})$, the lefthand side of the corresponding simultaneous equation in (E3) is $[\mathbf{Q}]_{hi} - [\mathbf{P}_A\mathbf{Q}\mathbf{P}_B']_{hi}$. And $[\mathbf{P}_A\mathbf{Q}\mathbf{P}_B']_{hi}$, i.e. $[\mathbf{P}_A]_{h.}\mathbf{Q}[\mathbf{P}_B']_{.i}$, is a homogeneous linear combination of the nonzero $\mathbf{Q}$-terms such that the coefficient of each $q_{(jk)}$ in $[\mathbf{P}_A\mathbf{Q}\mathbf{P}_B']_{hi}$ is simply $[\mathbf{P}_A]_{hj}[\mathbf{P}_B']_{ki}$. So equations (E3) can be written as a single matrix equation

$$(\underset{\sim}{I} - \underset{\sim}{S})\underset{\sim}{q} = \underset{\sim}{v} \tag{E5}$$

where $\underset{\sim}{S}$ is a matrix whose element in row (<u>hi</u>) and column (<u>jk</u>) is

$$[\underset{\sim}{S}]_{(hi)(jk)} =_{def} [\underset{\sim}{P}_A]_{hj}[\underset{\sim}{P}_B']_{ik}$$

and $\underset{\sim}{v}$ is a vector whose (<u>hi</u>)th element is

$$[\underset{\sim}{v}]_{(hi)} =_{def} [\underset{\sim}{M}_1]_{hi} - [\underset{\sim}{P}_A]_{h.}\underset{\sim}{M}_1[\underset{\sim}{P}_B']_{.i} \quad .$$

Unless $\underset{\sim}{S}$ is singular, solution of (E4) for the least-squares-optimal estimate of the nonzero $\underset{\sim}{Q}$-elements is then $\underset{\sim}{q} = (\underset{\sim}{I} - \underset{\sim}{S})^{-1}\underset{\sim}{v}$.

However, this simple solution for $\underset{\sim}{q}$ is likely to be complicated by equality constraints imposed on some of its free elements. For example, symmetry may be required of $\underset{\sim}{Q}$ even when some of its free elements are off-diagonal. Let the indices of $\underset{\sim}{q}$ be partitioned into blocks $\beta_1,\ldots,\beta_r$ such that the $\underset{\sim}{q}$-elements with indices in the same β_1 are constrained to be equal. Then by Lagrange-multiplier inclusion of these side conditions in the least-squares optimization it can easily be shown that the rows of (E5) with indices in the same block are replaced by the sum of these rows while of course in each row the previously distinct $\underset{\sim}{q}$-elements in each block are replaced by just one unknown. Specifically, (E5) reduces under equality-constraint blocks $\beta_1,\ldots,\beta_r$ to

$$(\underset{\sim}{I} - \underset{\sim}{S}_1)\underset{\sim}{q}_1 = \underset{\sim}{v}_1$$

wherein the <u>m</u>th element of $\underset{\sim}{v}_1$ and the <u>mn</u>th element of $\underset{\sim}{S}_1$ are respectively

$$[\underset{\sim}{v}_1]_m = \sum_{(hi)}^{\beta_m} [\underset{\sim}{v}]_{(hi)} \,, \qquad [\underset{\sim}{S}_1]_{mn} = \sum_{(hi)}^{\beta_m}\sum_{(jk)}^{\beta_n} [\underset{\sim}{S}]_{(hi)(jk)} \qquad (\underline{m},\underline{n} = 1,\ldots,\underline{r}) \,,$$

while $[\underset{\sim}{q}_1]_m$ is the free $\underset{\sim}{Q}$-element common to block β_m. ($\sum_{(hi)}^{\beta_m}$ here abbreviates summation over all the indices (<u>hi</u>) in block β_m.)

Note.

Matrix $\underset{\sim}{Q} - \underset{\sim}{P}_A\underset{\sim}{Q}\underset{\sim}{P}_B'$ in (E3) can be reorganized by the <u>vec</u> transformation as

$$(Q - P_A Q P_B')^c = Q^c - (P_A Q P_B')^c = (I - P_B \otimes P_A) Q^c \quad .$$

Each element of Q^c, and each row and each column of $I - P_B \otimes P_A$, corresponds to one pair of Q's row/column indices; and it is easily seen that the left-hand side of (E5) can be obtained by letting q be what remains of Q^c after deletion of terms not indexed in σ while S is the principal minor of $I - P_B \otimes P_A$ whose rows/columns are similarly picked out by σ. This construction makes clear the maximum number of free Q-elements for which (E5) has a unique solution: By definition, a symmetric matrix is a "projector" just in case all its nonzero eigenvalues are unity, one consequence of which is that if P is any $n \times n$ projector of rank r, $I - P$ is an $n \times n$ projector of rank $n - r$. Now, given that A (B) is of order $n_A \times n_A$ ($n_B \times n_B$) and has the left-inverse A^L (B^L) defined above, P_A (P_B) is an $n_A \times n_A$ ($n_B \times n_B$) projector whose rank is r_A (r_B); whence $P_B \otimes P_A$ is an $n_B n_A \times n_B n_A$ projector of rank $r_B r_A$, making $I - P_B \otimes P_A$ one of rank $n_B n_A - r_B r_A$. So long as the number of free Q-elements does not exceed $n_B n_A - r_B r_A$, it is thus always possible for σ to so-position them in Q that S in (E5) is nonsingular. Even so, because $I - P_B \otimes P_A$ does contain $r_B r_A$ linear dependencies, even a small principal minor S thereof can in some cases be singular if it is chosen infelicitously. What σ-selections are assured of avoiding this indeterminacy, we do not know.