Behavior Genetics, Vol. 26, No. 1, 1996

The Cholesky Approach: A Cautionary Note

John C. Loehlin[1]

Received 18 Jan. 1995—Final 27 July 1995

Attention is called to a common misinterpretation of a bivariate Cholesky analysis as if it were a common and specific factor analysis. It is suggested that an initial Cholesky behavior genetic analysis should often be transformed into a different form for interpretation. Formulas are provided for four transformations in the bivariate case.

KEY WORDS: Cholesky; triangular decomposition; simplex; multivariate behavior-genetic analysis.

Cholesky factoring, or triangular decomposition, is becoming a popular approach to multivariate behavior genetic problems (see, e.g., Neale and Cardon, 1992). In such a procedure, illustrated in Fig. 1 for three variables, the first latent variable, F_1, has effects on all the variables V_1 to V_3; the second, F_2, is uncorrelated with the first and has effects on the remaining variables V_2 and V_3; and the last, F_3, is specific to V_3. One use of the Cholesky procedure is in temporal contexts. For example, V_1 to V_3 might represent measurements of some variable at three successive times. In this case, F_1 would represent causes present at time 1 which affect the observed variable at time 1 and on subsequent occasions; F_2 would represent additional causes which arise by time 2 and whose effects are added to those of F_1 from time 2 on; and, finally, F_3 represents new causes at time 3 which affect only the last measurement, V_3.

However, a Cholesky decomposition can also represent a multivariate analysis of simultaneously measured variables *considered in some rationally defined order of priority*. In this case, F_1 is assigned the first priority, to explain V_1 and as much of V_2 and V_3 as it can. Then F_2, given second priority, explains what is left of V_2 and as much as it can of V_3. Finally, F_3 takes care of what is left of V_3.

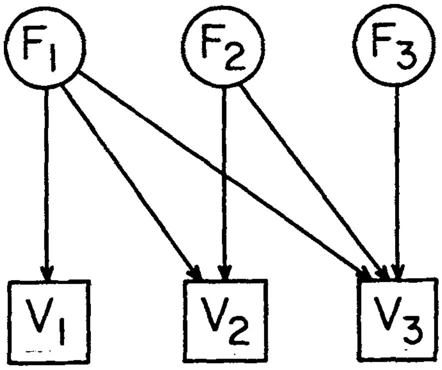

Fig. 1. Example of Cholesky decomposition. F_1, F_2, and F_3, Cholesky factors; V_1, V_2, and V_3, variables.

It should be emphasized that the explanation arrived at depends on the ordering—if we had considered the latent variables in the reverse order, F_3 would be a factor with paths to all the variables, and F_1 a residual. Only in the case of uncorrelated variables is the order of selection immaterial.

Any of the six possible orderings of the three latent variables in Fig. 1 explains the variance–covariance matrix among the variables equally well. Any solution can be transformed into any other at will. The justification of one over another as an explanation of the data depends entirely on the logic underlying the sequence in which the variables are considered. In the temporal case, there is

---

[1] Department of Psychology, University of Texas, Austin, Texas 78712. E-mail: loehlin@psy.utexas.edu.

a compelling underlying ordering, that of time sequence. In the general multivariate case, the ordering must be rationally justified. In many situations, rather than undertake such a justification, a user might prefer to transform an initial Cholesky solution into some alternative form that does not make this demand, along lines to be discussed shortly.

In typical behavior-genetic applications, each of the latent variables in a Cholesky analysis is decomposed into genetic and environmental components by obtaining data on the observed variables in an appropriate sample of twins, adoptees, or other relatives. An example of such a decomposition is given in Fig. 2, based on the familiar *A*, *C*, and *E* model of additive effects, common environmental effects, and unique environment plus error (e.g., Heath *et al.*, 1989). Other variations, such as solving for *D* instead of *C*, dividing the *C* or *E* term into more than one component, and so on, would follow the same principles.

A common misinterpretation of Cholesky analyses in the bivariate case is illustrated by Fig. 3. (For simplicity, all latent and observed variables are assumed to be in standard-score form.) The analysis on the left, labeled (i), is a Cholesky analysis, with variable *A* taken as primary. F_1 is the sole cause of variable *A* and a partial cause of *B*. F_2 is a residual, representing the variance in *B* independent of F_1 and *A*. The analysis on the right, (ii), is a common and specific factor analysis, with F_C a factor affecting both *A* and *B*, and F_A and F_B factors specific to *A* and *B*, respectively. The error is to speak of analysis (i) as though it were analysis (ii). An example would be to take a two-occasion measurement analyzed by a Cholesky model and describe the genetic component of F_1 as representing common genetic influences or genetic influences affecting the phenotype at both ages (e.g., Plomin *et al.*, 1994, p. 209). This latent variable does include such influences, so the statement is not entirely incorrect, but it also includes genetic influences that are *not* common to both ages but specific to time 1. Another example (Duffy *et al.*, 1994) involves two variables measured concurrently. Two indices of Type A personality were analyzed via a bivariate Cholesky analysis. The variance corresponding to *A* and *E* components of F_1 is described as "common additive genetic" and "common unshared environment"—despite the likely presence of a specific component in the former, and the almost certain presence of one in the latter. In still other cases of Cholesky analyses, although the description of the analysis is accurate, a full justification of the ordering of the variables is skimpy or absent (e.g., Truett *et al.*, 1992).

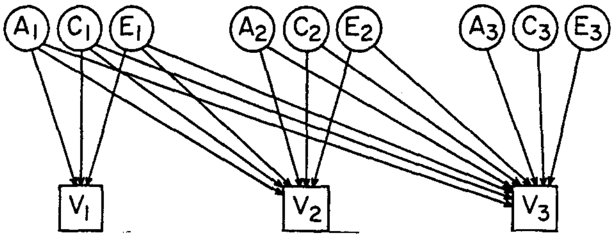

Fig. 2. Cholesky factors of Fig. 1 decomposed into additive genetic (*A*), common environmental (*C*), and unshared environmental (*E*) components.

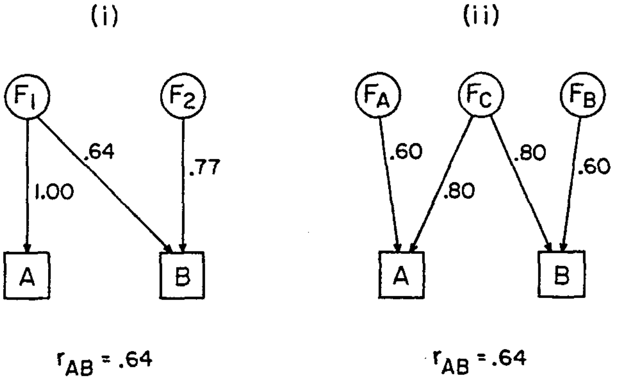

Fig. 3. Bivariate Cholesky model (i) compared to common and specific factors model (ii). *A* and *B*, observed variables. F_1 and F_2, Cholesky factors. F_C, common factor; F_A and F_B, specific factors.

In cases such as the above, the straightforward separation of influences into those that are shared and those that are specific would require a model of type (ii) rather than that of type (i). Type (ii) models are not generally solvable in the bivariate case, unless additional restrictions are imposed. On possibility is to require the paths from F_C to *A* and *B* to be equal. An example is provided by Schmitz *et al.* (1994). A model of type (ii) was fit to data on early childhood Internalizing and Externalizing behaviors by equating the paths to the two variables from the common *A*, *C*, and *E* factors. (Note that such a procedure normally makes sense only if the two variables are measured on comparable scales or are standardized.) With three or more variables,

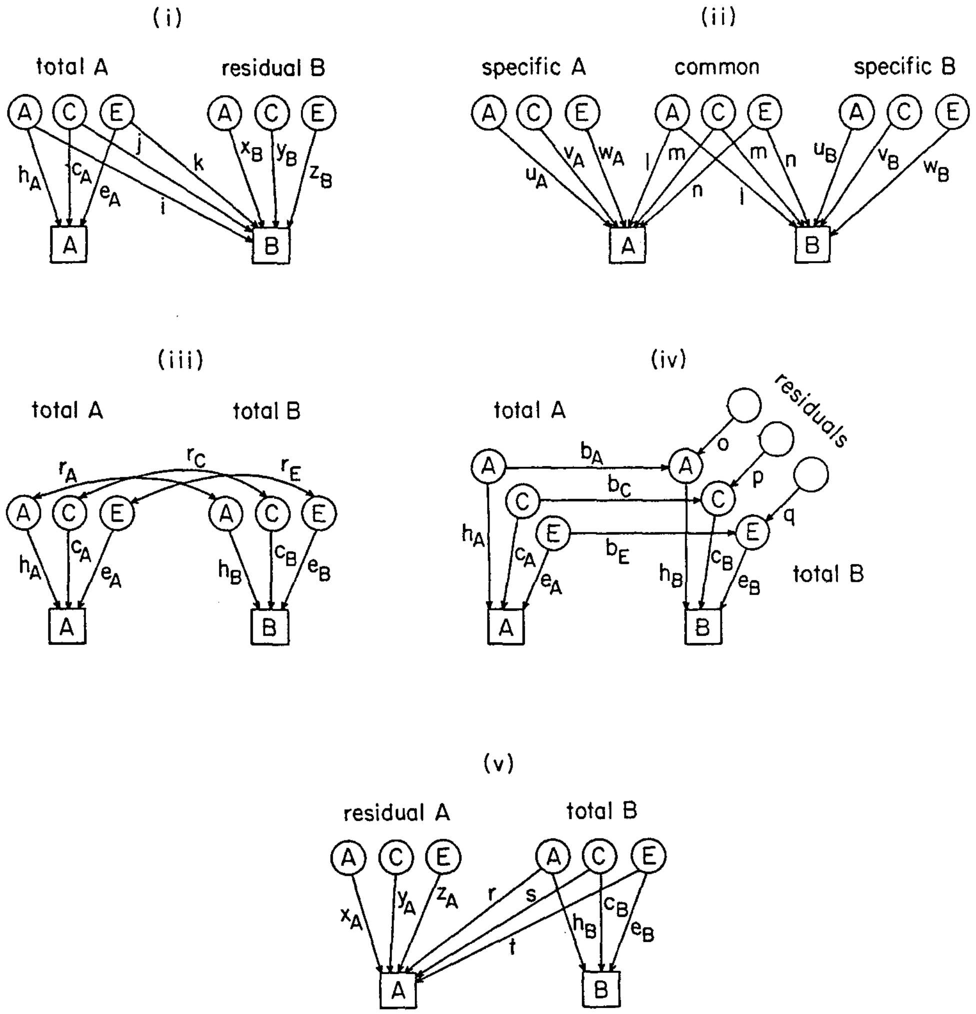

Fig. 4. Five alternative bivariate behavior genetic models: (i) Cholesky, (ii) common and specific factors, (iii) correlated factors, (iv) simplex, and (v) opposite Cholesky. In squares, A and B observed variables. In circles, A, C, and E, additive genetic, common environmental, and unshared environmental latent variables. Correlations: r_A, r_C, and r_E, additive genetic, common environmental, and unshared environmental correlations between corresponding components of A and B. Paths: h_A, c_A, e_A, h_B, c_B, and e_B, additive genetic, common environmental, and unshared environmental paths corresponding to univariate analyses of A and B; x, y, z, u, v, w, o, p, and q, residuals; b_A, b_C, and b_E, paths between additive genetic, common, environmental, and unshared environmental latent variables at times A and B. Remaining path labels arbitrary, for use in Table I equations.

model (ii) becomes the standard Spearman model of a general plus specific factors and, as such, is solvable except in special cases. Some recent behavior-genetic examples include Baker *et al.* (1991), Buhrich *et al.* (1991), and Petrill and Thompson (1993).

It would be easy to assume that at least the residuals for the second variable in type (i) and type (ii) models are the same, but they are not. Figure 3 contains numerical values derived from an assumed observed correlation of 0.64 between A and B. Note that the specific contribution to B differs depending on which model is being considered. If, as in (ii), a common factor equally correlated with A and B is assumed, 64% of B's variance is explained by it, and 36% is residual. But if model (i) is assumed, only 41% of B's variance is explained by F_1 and 59% is residual. Thus even an analysis of just the second variable into a part shared with the first and a part that is specific (cf. Duffy *et al.*, 1994, p. 473) can be misleading, if based on a Cholesky analysis. As always in latent

Table I. Transforming Bivariate Cholesky (i) to Other Models in Fig. 4 (All Latent and Observed Variables Assumed Standardized)

Model (ii), common and specific	Model (iii), correlated factors	Model (iv), simplex	Model (v), opposite Cholesky
$l = \sqrt{ih_A}$ $m = \sqrt{jc_A}$ $n = \sqrt{ke_A}$	$h_B = \sqrt{i^2 + x_B^2}$ $c_B = \sqrt{j^2 + y_B^2}$ $e_B = \sqrt{k^2 + z_B^2}$	$\left.\begin{matrix} h_B \\ c_B \\ e_B \end{matrix}\right\}$ same as (iii)	$\left.\begin{matrix} h_B \\ c_B \\ e_B \end{matrix}\right\}$ same as (iii)
$u_A = \sqrt{h_A^2 - l^2}$ $v_A = \sqrt{c_A^2 - m^2}$ $w_A = \sqrt{e_A^2 - n^2}$	$r_A = i/h_B$ $r_C = j/c_B$ $r_E = k/e_B$	$b_A = i/h_B$ $b_C = j/c_B$ $b_E = k/e_B$	$r = ih_A/h_B$ $s = jc_A/c_B$ $t = ke_A/e_B$
$u_B = \sqrt{i^2 + x_B^2 - l^2}$ $v_B = \sqrt{j^2 + y_B^2 - m^2}$ $w_B = \sqrt{k^2 + z_B^2 - n^2}$	$\left.\begin{matrix} h_A \\ c_A \\ e_A \end{matrix}\right\}$ same as (i)	$\left.\begin{matrix} h_A \\ c_A \\ e_A \end{matrix}\right\}$ same as (i)	$x_A = \sqrt{h_A^2 - r^2}$ $y_A = \sqrt{c_A^2 - s^2}$ $z_A = \sqrt{e_A^2 - t^2}$
		$o = \sqrt{1 - b_A^2}$ $p = \sqrt{1 - b_C^2}$ $q = \sqrt{1 - b_E^2}$	

variable modeling, care must be taken that the language used in the interpretation corresponds to the analysis that was actually carried out.

One possible way to proceed in many cases is to obtain an initial Cholesky solution and transform it into the desired form for interpretation. Alternatively, one could fit the desired solution directly, but the Cholesky is easy to program and solve, and provides a reasonable starting point if one wants to look at several alternatives. [Some mathematical virtues of the Cholesky solution are discussed by Neale and Cardon (1992).]

Figure 4 represents five solutions for the two-variable case. Models (i) and (ii) are the *A,C,E* versions of the diagrams in Fig. 3—the Cholesky and a common and specific factor model. Model (iii) is what Neale and Cardon (1992, p. 270) call the "correlated factors" model. In this, each variable is separately decomposed into its genetic and environmental components, and the correlations of these across variables are estimated. Thus, for example, variables *A* and *B* might each be highly heritable but be influenced by different genes (r_A = 0). At the same time, it might be that the few environmental events that do affect *A* and *B* tend to influence both—whether these events are shared by family members or unique to an individual (r_C and r_E large). Model (iv) is a simplex (e.g., Boomsma and Molenaar, 1987), in which causes present at time *A* partially persist until time *B* (the regressions b_A, etc.), at which time new causes may enter (the residuals). Finally, model (v) is a second Cholesky analysis taking trait *B* as primary—one might sometimes want to consider how different the effects are of making the two extreme assumptions about the causal priority of *A* and *B*.

Table I gives simple formulas for transforming the bivariate Cholesky solution in (i) into any of the other four. Those for the correlated factors solution (iii) are given in a slightly different form by Neale and Cardon (1992, p. 272). It should be emphasized that these are all alternative ways of looking at the same facts. All produce exactly the same implied variance–covariance matrix and, hence, the identical overall goodness of fit. However, they represent different causal models of what is going on, and the differences in the values of their paths reflects this fact.

Unless a strict logical priority holds between variable *A* and variable *B*, models (ii) and (iii) are likely to be the most readily interpretable. Model (i) should remain useful for dealing with a single variable measured on two occasions, although (iv), (ii), and (iii) can also be employed in this context—the last as reflecting genetic and environmental correlations over time (e.g., Plomin and DeFries, 1981). It should be noted that a given transformation may not always be possible. The application of a formula can sometimes lead to impossible results, such as a negative square root or a negative residual variance. This simply means that a model of the desired form will not do the job. However, a variant of the solution constrained to avoid the undesirable feature may still be acceptable. This is probably best achieved by fitting the desired variant directly, with the necessary constraints imposed. It should also be noted that simple transformational formulas of the present kind do not necessarily generalize easily to larger numbers of variables, although the interpretational issues apply to such cases as well.

In summary, the approach to multivariate behavior-genetic analysis via Cholesky decomposition can be a useful first step, but one that should be interpreted with care and, in many cases, not left as a final solution. One can make an analogy to the use of principal components in factor analysis: principal components are computationally convenient and have many attractive mathematical properties, but they are usually not interpreted without further transformation. In fact, it is of some interest that the Cholesky procedure was itself at one time in use in factor analysis as a method of initial factoring, under the rubric of the *diagonal method* (Harman, 1976, p. 101). The further development and use of rotational or other transformational procedures for following up initial solutions in behavior genetic analysis would seem to be a worthwhile goal, as pointed out some years ago by Crawford and DeFries (1978).

ACKNOWLEDGMENTS

I appreciate the helpful comments of two anonymous referees.

REFERENCES

Baker, L. A., Vernon, P. A., and Ho, H.-Z. (1991). The genetic correlation between intelligence and speed of information processing. *Behav. Genet.* **21**:351–367.

Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**:111–123.

Buhrich, N., Bailey, J. M., and Martin, N. G. (1991). Sexual orientation, sexual identity, and sex-dimorphic behaviors in male twins. *Behav. Genet.* **21**:75–96.

Crawford, C. B., and DeFries, J. C. (1978). Factor analysis of genetic and environmental correlation matrices. *Multivar. Behav. Res.* **13**:297–318.

Duffy, D. L., Manicavasagar, V., O'Connell, D., Silove, D., Tennant, C., and Langelludecke, P. (1994). Type A personality in Australian twins. *Behav. Genet.* **24**:469–475.

Harman, H. H. (1976). *Modern Factor Analysis*, 3rd ed., University of Chicago Press, Chicago.

Heath, A. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., and Fulker, D. W. (1989). Testing structural equation models for twin data using LISREL. *Behav. Genet.* **19**:9–35.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*, Kluwer, Dordrecht, The Netherlands.

Petrill, S. A., and Thompson, L. A. (1993). The phenotypic and genetic relationships among measures of cognitive ability, temperament, and scholastic achievement. *Behav. Genet.* **23**:511–518.

Plomin, R., and DeFries, J. C. (1981). Multivariate behavioral genetics and development: Twin studies. In Gedda, L., Parisi, P., and Nance, W. E. (eds.), *Twin Research 3. Part B: Intelligence, Personality, and Development*, Liss, New York, pp. 25–33.

Plomin, R. Pedersen, N. L., Lichtenstein, P., and McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behav. Genet.* **24**:207–215.

Schmitz, S., Cherny, S. S., Fulker, D. W., and Mrazek, D. A. (1994). Genetic and environmental influences on early childhood behavior. *Behav. Genet.* **24**:25–34.

Truett, K. R., Eaves, L. J., Meyer, J. M., Heath, A. C., and Martin, N. G. (1992). Religion and education as mediators of attitudes: A multivariate analysis. *Behav. Genet.* **22**: 43–62.

Edited by David W. Fulker