LISREL analysis of twin data with structured means

Dolan, C.V.; Molenaar, P.C.M.; Boomsma, D.I.

published in
Behavior Genetics
1989

DOI (link to publisher)
10.1007/BF01065883

document version
Publisher's PDF, also known as Version of record

Behavior Genetics, Vol. 19, No. 1, 1989

LISREL Analysis of Twin Data with Structured Means

C. V. Dolan,[1] P. C. M. Molenaar,[1] and D. I. Boomsma[2]

A method is introduced to test the hypothesis that both the phenotypic means and the phenotypic covariances can be modeled with the same common genetic and environmental factors. LISREL can be used to implement the method. An illustration is given with simulated twin data.

KEY WORDS: covariance structure; means structure; augmented moment matrices; twin data; LISREL.

INTRODUCTION

Behavioral genetic research generally focuses on the contribution of heredity and the environment to individual differences. It has been maintained that this approach is onesided because it precludes causal modeling of phenotypic means. Notably, McCall (1981) has argued that the structural models of both the mean and the covariance are equally interesting and, in fact, complementary pieces of information in understanding the ontogenesis of a behavioral trait.

For instance, the structure of the means is relevant in the comparison of age cohort samples with regard to the effects of environment and heredity (Scarr-Salapatek, 1976). The finding that heredity makes a large contribution to the variance does not necessarily have a bearing on an observed difference in means between cohorts. It is not possible to interpret such a finding in terms of genetical and environmental effects

This work was aided by Nato Grant 86/0823 and grants from the Belgian National Fund, the State University of Gent, and the Catholic University of Leuven.

[1] University of Amsterdam, Department of Psychology, Weesperplein 8, 1018 XA Amsterdam, The Netherlands.

[2] Free University, Department of Experimental Psychology, De Boelelaan 1115, 1081 HV Amsterdam, The Netherlands.

0001-8244/89/0100-0051$06.00/0

without further knowledge of the relationship between the structures of covariance and means.

Causal models have been used to study the covariance structure and the structure of the mean simultaneously (e.g., Sörbom, 1982; McArdle and Epstein, 1987). Such models have been applied to the analysis of twin data by McArdle (1986) within his reticular action model (RAM) (McArdle and McDonald, 1984). In this approach, McArdle does not estimate the means of the genetical and environmental factors (second-order factors in the multivariate application of the RAM), but of a first-order psychometric factor. The structure of the observed means is therefore not related directly to the biometrical factors.

In this paper, we suggest a method which involves assessing the influence of the biometrical factors upon the observed means within the context of the covariance analysis of twin data suggested by Martin and Eaves (1977). The objective of this method is to test explicitly the hypothesis that the structures of the phenotypic means and covariances can be modeled by the same common environmental and genetical factors. When this is found to be the case, it implies that the same genetic and environmental effects account for both the individual differences and the phenotypic means.

The method, which can be carried out with the program LISREL VI (Jöreskog and Sörbom, 1986), involves comparing the overall goodness of fit obtained from the analysis of covariance with the goodness of fit obtained from the simultaneous analysis of phenotypic means and covariance. It does, however, require a minimum number of observed variables equaling one plus the number of factors in the twin model of which means are estimated.

An illustration is given with simulated data.

LISREL ANALYSIS OF TWIN DATA

The factor analytic approach to the genetical analysis of covariance structure has been explained by Martin and Eaves (1977) and the implementation of the LISREL program to this end has been discussed elsewhere (Boomsma and Molenaar, 1986; see also this issue). These subjects are, therefore, dealt with briefly.

The analysis of the monozygotic (MZ) and dizygotic (DZ) covariance structures amounts to a confirmatory factor analysis in which the adequacy of a model, which contains environmental and genetic factors, as an explanation of the observed covariance structure is tested. In the multivariate case, the variables that are observed in Twin 1 and Twin 2 of the MZ and DZ samples are constrained to load on the common genetic

and environmental factors associated with each twin. Specific environmental and genetical factors can be introduced to accommodate the variance that is not explained by the communal part of the model. The genetical relatedness is expressed in the fixed correlation of the factors: the additive genetic factors in the MZ group correlate perfectly; those in the DZ group have an expected correlation of .5. The loadings of the observed variables on the factors are constrained to be equal between Twin 1 and Twin 2 and across zygosities. These loadings are generally estimated by maximum likelihood. The model for the covariance structure can be expressed as follows in the LISREL notation:

$$\sum_i = \Lambda\Psi_i\Lambda', \qquad i = \text{mz, dz.} \tag{1}$$

If n is the number of observed variables and p is the number of factors, lambda is a $(n \times p)$ matrix containing the loadings of observed variables on latent factors and psi is a $(p \times p)$ correlation matrix of the latent factors.

The objective of analyzing twin data with LISREL is to find a model which fits the data and subsequently to assess the contributions of the factors to the observed variances. The overall fit of the model is tested by chi-square and the significance of individual parameter estimates is judged against their standard errors or by dropping them from the model (see Neale *et al.*, 1989).

In this approach, the expected values of the factors are assumed to equal zero (e.g., Boomsma and Molenaar, 1986) and the expected values of the observed variables are removed by summarizing the data in covariance matrices.

LISREL ANALYSIS OF THE STRUCTURE OF MEANS

It is possible to include the means of both the observed and the latent variables in the analysis of twin data to test the dependence of the structures of means and covariance. This can be achieved by allowing certain factors to influence other factors selectively. In LISREL this involves utilizing the beta matrix. The LISREL model then becomes

$$\sum_i = \Lambda(I - B)^{-1}\Psi_i(I - B')^{-1}\Lambda', \qquad i = \text{mz, dz.} \tag{2}$$

The data are summarized in augmented moment (AM) matrices, instead of covariance matrices. The AM matrix is the matrix of raw moments (i.e., calculated without mean correction) after a dummy variable which equals one has been added to each case. In the present application, a case is a twin pair. The dummy variable, which is added to both the MZ and the DZ samples, ensures that the last, n + 1th, row of the input matrices

contains the observed means of Twin 1 and Twin 2. The last, n + 1th, element in this row contains the average cross product of the dummy variable with itself, i.e., one. LISREL has facilities for computing the AM matrices from either raw data or summary statistics.

The model that is tested in the present application contains the common factors of the covariance model, e.g., an individual environmental factor (E), an additive genetic factor (G), and an additional factor (D) to accommodate the dummy variable (the number of factors is now p + 1). Unique factors may be introduced to accommodate specific variances but they are assumed not to contribute to the phenotypic means. The observed variables are constrained to load on the environmental and genetic factors as outlined above. The dummy variable is fixed to load on the additional factor, D, with a loading equal to 1.0. The beta matrix is used to estimate the influence of the additional factor on the common genetic and environmental factors by freeing the relevant elements of this matrix.

An illustrative path diagram of this model is given in Fig. 1. Expressing the model for Fig. 1 more elaborately clarifies how the observed means and covariances are modeled simultaneously. Figure 2 contains

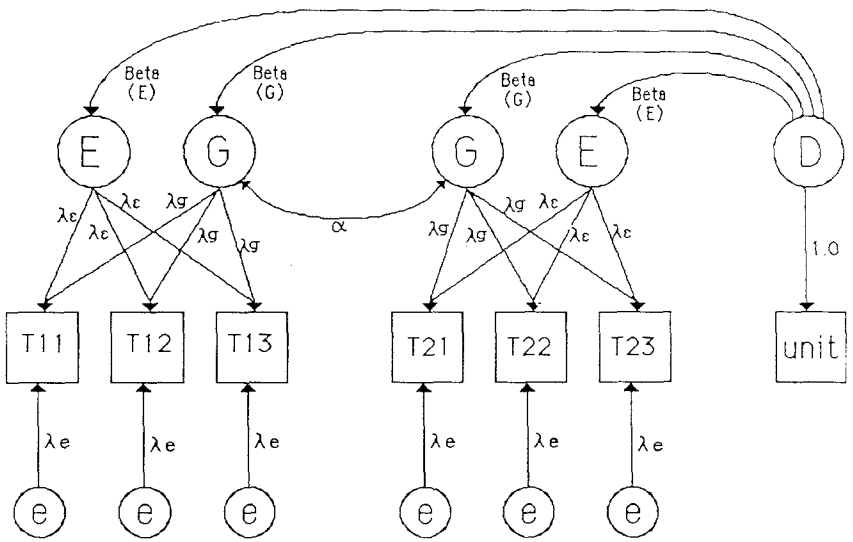

Fig. 1. An E,G twin model with structured means. The observed variables (T's) are constrained to load upon the additive genetic (G) and environmental factors (E). Unit is the dummy variable and D is the additional factor. The e's are measurement errors or unique environmental influences. Alpha is the correlation between the additive genetic factors. For MZs α = 1.0; for DZs α = .5. The beta coefficients are estimates of the biometrical factor means which determine the observed means through the factor loadings.

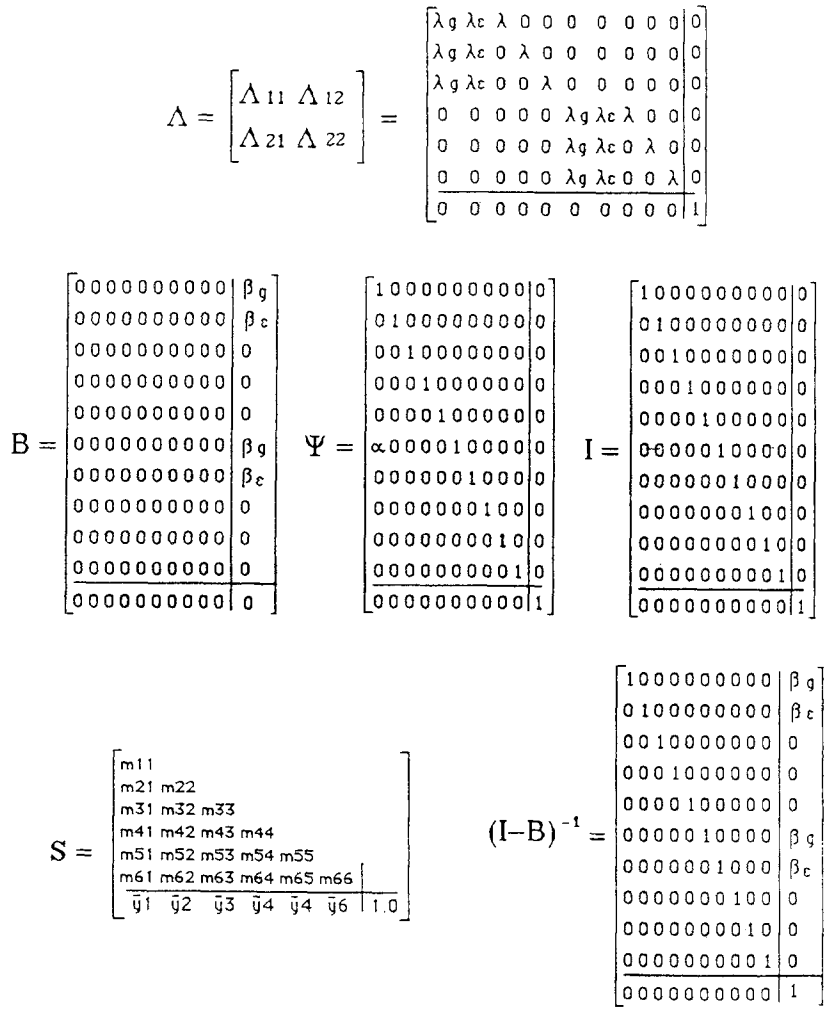

Fig. 2. The partitioned matrices of Eq. (2). All matrices are decomposed in a manner similar to the Λ matrix, i.e., into four indexed submatrices. Alpha in the Ψ matrix represents the correlation between the additive genetic factors of Twin 1 and Twin 2.

the matrices of Eq. (2) in their full form. Each matrix is partitioned into submatrices that accommodate the covariances structure, the means structure, and the presence of the dummy variable. Expressed in terms of the partitioned matrices the right-hand side of Eq. (2) can be written as

$$\Sigma = \left[\begin{array}{c|c} \Lambda_{11}\Psi_{11}\Lambda'_{11} + \Lambda_{11}(I - B)\bar{}^{1}_{12}(I - B')\bar{}^{1}_{12}\Lambda'_{11} & \Lambda_{11}(I - B)\bar{}^{1}_{12} \\ \hline (I - B')\bar{}^{1}_{12}\Lambda'_{11} & 1 \end{array}\right]. \tag{3}$$

Here, $\Lambda_{11}\Psi_{11}\Lambda'_{11} + \Lambda_{11}(I - B)\bar{}^{1}_{12}(I - B')\bar{}^{1}_{12}\Lambda'_{11}$ represents the matrix of moments of the observed variables, and $\Lambda_{11}(I - B)\bar{}^{1}_{12}$ is a vector of observed means. Comparing Eq. (3) to S, the sample AM matrix in Fig. 2 reveals how the raw moments are expressed in terms of the model's parameters. Note that the average cross product between two variables equals the sum of the covariance of the variables and the product of their mean values.

When common factors account for both the covariance and the means structure, the estimates of the elements in beta can be interpreted as the expected values of the common factors, that is,

$$\begin{aligned} E[E] &= \beta_e E[D] = \beta_e, \\ E[G] &= \beta_g E[D] = \beta_g, \end{aligned} \tag{4}$$

where $E[D]$ denotes the expected value of the additional factor and β_g and β_e are the loadings of the biometrical factors on the additional factors. The expected values of the common factor reduce to the beta weights because the expected value of the additional factor equals one.

The expected values of the observed variables are then

$$E[T] = \lambda_e\beta_e + \lambda_g\beta_g. \tag{5}$$

The hypothesis that the common biometrical factors (E and G) account for the structures of both means and covariances can be tested by fitting the model both with and without the structured means. When this is the case, the parameter estimates in lambda obtained from the analysis of covariance matrices should equal those obtained from the analysis of augmented moment matrices, and the overall fit (as indicated by the chi-square statistic) should be approximately as good as the overall fit obtained from the analyses of covariance matrices.

A significant decline in the chi-square when the means are introduced into the analysis indicates that the common factors that account for the covariance cannot account for the mean structure.

The introduction of structured means into the analysis of twin data as suggested here requires a minimum number of observed variables. The number must exceed the number of common factors by one in order to ensure the identification of the factor means. In other words, the number of equations associating the observed means with the common factor means must exceed the number of factor means that are estimated.

Table I. True Parameter Values

	G	E	e	Phenotypic mean
Phenotype 1	2.50	7.10	1.0	76.40
Phenotype 2	4.25	2.90	1.0	47.35
Phenotype 3	3.10	5.20	1.0	62.30
	$E[G] = 5.0$	$E[E] = 9.0$		

The suggested procedure, then, is to find a model that fits one's data by analyzing the covariance matrices and, subsequently, fit the model plus the means structure to the AM matrices. If the chi-square indicates that the latter model still fits, the common genetic and environmental factors that contribute to the individual differences also contribute to the phenotypic means.

An illustration of this method is given with simulated data.

ILLUSTRATION

For 250 MZ and DZ twin pairs three-variate phenotypes were generated with the IMSL subroutine GGNSM (IMSL, Inc., 1979) according to the *E, G* (individual environment and additive genetic effects) model depicted in Fig. 1. As the means of both E and G will be estimated, a minimum of three observed variables is required. The parameter values are given in Table I along with the expected values of the factors and the phenotypes. The 6×6 covariance matrices and the 7×7 AM matrices were calculated for the MZ and DZ twins. The results the analysis of the covariance structure are given in Table II. The true parameters are recovered nicely: the model, judging by the chi-square, fits like a glove.

Table II. Analysis of Covariance Structure

	Parameter estimate (SE in parentheses)		
	G	E	e
Phenotype 1	2.307 (.262)	7.061 (.168)	1.041 (.128)
Phenotype 2	4.197 (.156)	2.897 (.100)	.958 (.048)
Phenotype 3	2.932 (.204)	5.197 (.130)	.955 (.059)
	$E[G] = .0$	$E[E] = .0$	
	$\chi^2_{33} = 16.18$	$p = .99$	

Table III. Analysis of AM Matrices: Identical Structures of Covariances and Means

	Parameter estimate (SE in parentheses)			Estimated phenotypic mean
	G	*E*	*e*	
Phenotype 1	2.303 (.261)	7.048 (.166)	1.041 (.118)	76.26
Phenotype 2	4.191 (.155)	2.892 (.099)	.956 (.048)	47.30
Phenotype 3	2.927 (.203)	5.189 (.128)	.953 (.054)	62.22
	$E[G]$ = 4.93 (.232)	$E[E]$ = 9.20 (.281)		
	χ^2_{43} = 18.73	p = .99		

Subsequently, the AM matrices were used as input. It is perhaps worthwhile to dwell on the actual implementation of the program. First, on the DATA card in LISREL the fact that AM matrices are to be analyzed is specified by the statement MA = AM. The number of input variables is given as the number of real observed variables, so here we have NI = 6. LISREL will compute the AM matrices from the raw data or from any summary of the data as long as the covariances and means can be derived. On the MODEL card, the number of variables should now include the dummy unit variable, so NY = 7. The number of factors is simply the number of factors in the covariance structure model plus the additional factor, *D*, for the dummy variable, i.e., NE = 10 + 1. LY contains the factor loadings where it is important that the unit dummy variable loads exclusively on the additional factor, *D*, with a fixed loading of 1.0. BE contains the regressions of the biometrical factors upon the factor *D*. Finally, PS is the correlation matrix of the factors. The loadings of the observed variables on the factors are constrained to be equal across the DZ and MZ groups, as are the loadings of the environmental and genetic factors on the additional factor [see Appendix (Fig. A1) for the LISREL setup].

The results of the analysis of the model incorporating the structures of both the means and the variance are given in Table III. The parameter estimates of the factor loadings are almost identical to those obtained from the analysis of covariance matrices (Table II) and are close to the true values. The expected values of the common biometrical factors are simply the parameter estimates of the freed beta elements. These, too, do not diverge greatly from the true values. Judging the chi-square by the one obtained from the analysis of covariance matrices, it would be safe to conclude that the structures of the phenotypic means and covariances can be accounted for by the same common biometrical factors. It should be pointed out that the average cross product of the dummy variable with

itself in the input AM matrices cannot be taken as an independent statistic and therefore does not contribute to the degrees of freedom. Because this is not recognized by the LISREL program as it is used here, it is necessary to subtract one degree of freedom for each group from the degrees of freedom given in the LISREL output and calculate anew the probability level of the obtained chi-square.

A second data set was generated according to the model in Fig. I but with independent structures of means and covariances. The covariances structure was analyzed first. Here the fit was good, with a χ^2_{33} of 41.97 ($p = .135$). Subsequently, the model involving the structures of both the means and the covariances was tested. The fit of the model was found to be extremely poor ($\chi^2_{43} = 1823.32$, $p = .00$), indicating that the common factors cannot account for the structure of the observed means.

DISCUSSION

The objective of the method discussed in this paper is to test the hypothesis that both the phenotypic means and the phenotypic covariance can be modeled by the same common biometrical factors. Implementation of the method requires a minimum number of observed variables that depends on the number of common factors in the model for the covariance structure. The minimum number of observed variables must equal the number of factors in the communal part of the model plus one. As each observed mean is expressed as the weighted contributions of the means of the common biometrical factors [see Eq. (5)], a condition for the identification of the observed means is that the number of observed variables exceeds the number of these factors. Note that a similar situation is found in most textbooks where the identification of the population mean of a trait is achieved by fixing the expected value of the environmental factor to equal zero (Falconer, 1960, p. 112).

The investigation into the relationship between the structure of the means and that of the covariances could easily be extended to involve the comparison of different samples of twins such as male and female twins or the same sample at different points in time. This would involve following the procedure that is used in fitting a sex-limitation model (e.g., see Eaves *et al.*, 1978), that is, comparing a model in which the factor loadings (both lambdas and betas) are constrained to be equal across sexes with the model in which they are free to differ.

In conclusion, the method discussed in this paper has the potential to provide a fuller picture of the influence of heredity and environment on behavior.

APPENDIX

```
three variate mz twins - model E,G - identical structures means covar.
da ng=2 ni=6 no=250 ma=am
la
*
't11' 't12' 't13' 't21' 't22' 't23'
cm
*
55.4460
29.8899 26.1734
42.5224 26.7325 35.6698
 5.0615  9.9431  6.7500 59.0315
 9.8208 17.4062 12.3588 32.0683 27.8959
 6.9368 12.4777  8.7804 46.5600 29.1923 39.4146
me
(6f7.4)
76.499 47.506 62.393 76.149 47.331 62.133
mo ny=7 nx=0 ne=11 nk=0 ly=fu,fr ps=sy,fi te=ze be=fu,fi
pa ly
*
1 1 1 0 0 0 0 0 0 0 0
1 1 0 1 0 0 0 0 0 0 0
1 1 0 0 1 0 0 0 0 0 0
0 0 0 0 0 1 1 1 0 0 0
0 0 0 0 0 1 1 0 1 0 0
0 0 0 0 0 1 1 0 0 1 0
0 0 0 0 0 0 0 0 0 0 1
eq ly(1,3) ly(4,8)
eq ly(1,2) ly(4,7)
eq ly(1,1) ly(4,6)
eq ly(2,1) ly(5,6)
eq ly(2,2) ly(5,7)
eq ly(2,4) ly(5,9)
eq ly(3,1) ly(6,6)
eq ly(3,2) ly(6,7)
```

Fig. A1. Appendix. LISREL input to test the equivalence of means and covariances structures.

```
eq ly(3,5) ly(6,10)
fi ly(7,11)
va 1.0 ly(7,11)
ma ps
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 1 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 1
fr be(1,11) be(2,11) be(6,11) be(7,11)
eq be(1,11) be(6,11)
eq be(2,11) be(7,11)
st 5.0 all
ou se rs ns
three variate dz twins - model E1,G - identical structures means covar.
da no=250 ma=am
cm
*
53.5842
28.8904 26.9180
41.1134 26.6056 34.7054
 2.4443  5.8008  3.2128 57.1301
 5.9953 10.4549  7.2673 29.8899 27.0452
 3.2634  6.8865  4.2622 43.7938 27.1483 36.4573
me
(6f7.4)
76.100 47.080 62.094 76.331 47.350 62.312
mo ly=in ps=sy,fi te=in be=in
ma ps
1
```

Fig. A1. (*Continued*)

```
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 5 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 1
st 5.0 all
ou se rs ns
```

Fig. A1. ***(Continued)***

REFERENCES

Boomsma, D.I., and Molenaar, P. C. M. (1986). Using LISREL to analyze genetic and environment covariance structures. *Behav. Genet.* **16:**237–250.

Eaves, L. J., Last, K. A., Young, P. A., and Martin, N. G. (1978). Model fitting approaches to the analysis of human behavior. *Heredity* **41:**249–320.

Falconer, D. S. (1960). *Introduction to Quantitative Genetics*, Oliver and Boyd, Edinburgh.

IMSL, Inc. (1979). *IMSL Library Reference Manuel Edition 7*, IMSL, Inc., Houston, Tex.

Jöreskog, K. G., and Sörbom, D. (1986). *LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*, National Educational Resources, Chicago.

Martin, N. G., and Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity* **38:**79–93.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behav. Genet.* **16:**163–200.

McArdle, J. J., and Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Dev.* **58:**110–133.

McArdle, J. J., and McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *Br. J. Math. Stat. Psychol.* **37:**234–251.

McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Dev.* **52:**1–12.

Scarr-Salapatek, S. (1976). An evolutionary perspective of infant intelligence: Species patterns and individual variations. In Lewis, M. (ed.), *Origins of Intelligence*, Plenum, New York, pp. 165–197.

Sörbom, D. (1982). Structural equation models with structured means. In Jöreskog, K. G., and Wold, H. (eds.), *Systems Under Indirect Observation*, Elsevier North-Holland, New York.