De Modellen-sectie

A STATISTICAL MODEL FOR THE ANALYSIS OF COVARIANCE WITH FALLIBLE COVARIATES

by

Dag Sörbom

Abstract

Analysis of covariance has been used extensively in experimental and quasi-experimental studies to remove preexisting differences among treatment groups in order to even out these differences and to arrive at a more powerful analysis. However, as has been recognized by several recent authors, when the measurement of the covariate involves error, the adjustments may introduce severe biases into the analysis.

In this paper a general statistical model is described, which utilizes information from several covariates to make the proper adjustments. The model can handle any number of covariates, criterion variables, and any number of treatment groups simultaneously. By the use of the model and the estimation method described in the paper a wide variety of hypotheses concerning the parameters of the model can be tested by means of a large sample likelihood ratio test.

1. Introduction

The use of analysis of covariance (ANCOVA) in the educational and psychological sciences has been questioned for several reasons (see e.g. Lord, 1960, 1963, and 1967; Elashoff, 1969; Cronbach et al., 1976). In this paper a statistical model is presented which covers the following instances, when the usual ANCOVA fails to make the proper adjustment:

a. The covariate is not error-free and the reliability of the covariate is not exactly known for the application at hand.

b. There exists a treatment-covariate interaction effect, i.e. the slopes in the regression of criterion variable on covariate are unequal in different treatment groups.

c. The variances of the criterion variable given the covariate are unequal.

d. The measure for the criterion variable is fallible.

In this paper the estimation of the parameters of the model is described and a general method for testing hypotheses about the parameters is given. A special case of the model is treated in Sörbom (1974) and this has been used in the field of measurement of change (Sörbom, 1976). A similar approach has been given by Keesling and Wiley (1975), but their estimates are not fully efficient, since they use a two-stage estimation method, and with their model it is not possible to utilize prior information about similarities among groups to obtain comparable parameter estimates.

It should be noted that nonrandom assignment of cases to treatment cannot be handled by the model. No doubt, this is one of the main deficiencies of ANCOVA, but as an often cited passage of Lord (1963) firmly expresses it:

"If the individuals are not assigned to the treatment at random, then it is not helpful to demonstrate statistically that the groups after treatment show more difference than would be expected by random assignment - unless, of course, the experimenter has special information showing that the non-random assignment was nevertheless random in effect. If, as often happens, randomized assignment is impossible then there is often no way to determine what is the appropriate adjustment to be made for initial differences, and hence often no way to show convincingly by statistical manipulations that one treatment is better than another."

The effects of using a covariate that does not account for all pre-existing differences among the treatment groups cannot be handled by the statistical model, either.

The paper is concerned with a situation when there is data available from a number of groups of individuals. The groups are supposed to be representative samples from some populations, and we are interested in estimating the parameters describing these populations, and in studying differences among these parameters across populations.

The main objective of the paper is to discuss a statistical model, its identification and an estimation method which makes it possible to estimate the regression of true criterion score (or universe score, see Cronbach et al., 1972) on true covariate score for a number of treatment groups. Using this model it is possible to test the assumption of equal slopes in these regressions among the groups, as well as assumptions of equal error variances, equal true score variances, whether the measurements are parallel or tau-equivalent (Lord and Novick, 1968) and so on.

The use of the model is illustrated by analysis of two small sets of data. Also, a strategy for modifying the model when it does not fit data sufficiently well is given.

2. Analysis of covariance, a brief introduction

This section gives a very brief introduction to the general aspects of ANCOVA and introduces the notation to be used in the sections to follow. The main reason for the use of ANCOVA is that one is interested in studying the effect of a number of treatments. To this end, a number of cases, $N^{(g)}$, are randomly assigned to the g:th treatment group. The main focus is on the expected values of a criterion variable, $\eta^{(g)}$. We assume that there exists a covariate, $\xi^{(g)}$, a variable which to some extent accounts for preexisting differences among the groups. Thus, it is desirable to eliminate the effect of the covariate, which, in addition, will result in a more powerful analysis (see e.g. Bock, 1975). This is done by considering the linear regressions

$$\eta^{(g)} = \alpha^{(g)} + \gamma^{(g)} \xi^{(g)} + \zeta^{(g)} \tag{1}$$

for each group $g=1,2,\ldots,G$, where $\zeta^{(g)}$ denotes the error of the regression. The main focus is now on the intercept, $\alpha^{(g)}$. However, often the covariate $\xi^{(g)}$ cannot be measured without error, and then it is a wellknown fact (see e.g. Lord, 1960) that the errors can cause serious bias in the estimates of $\alpha^{(g)}$. Suppose that we have two variables, $x_1^{(g)}$ and $x_2^{(g)}$, that give us information of $\xi^{(g)}$, and that these measure $\xi^{(g)}$ in the following sense

$$\begin{aligned} x_1^{(g)} &= \mu_1 + \lambda_1 \xi^{(g)} + \epsilon_1^{(g)} \\ x_2^{(g)} &= \mu_2 + \lambda_2 \xi^{(g)} + \epsilon_2^{(g)} \end{aligned} \tag{2}$$

where $\epsilon_1^{(g)}$ and $\epsilon_2^{(g)}$ represent measurement errors in $x_1^{(g)}$ and $x_2^{(g)}$, respectively. This will be referred to as the measurement model (cf. Keesling and Wiley, 1975). In (2), μ_i is an arbitrary location parameter describing the level of the observable variable $x_i^{(g)}$, $i=1,2$. The inclusion of these parameters in the model implies that the other parameters are unaffected by adding or subtracting a constant to the observed variables. This feature of the model is important in several instances, since it is often not possible to determine a natural origin for the observed measures. Each λ_i is a parameter describing the scale of the variable $x_i^{(g)}$ as compared with the scale of other variables measuring the same $\xi^{(g)}$. It is seen from (2) that the variance of $\xi^{(g)}$ and the two λ-parameters are not identified, since we can multiply $\xi^{(g)}$ by a constant and divide the λ:s by the same constant, without changing the observed measurements at all. Therefore, in the following, λ_1 is fixed equal to 1. The parameters μ_i and λ_i have no superscript, (g), because in general the same variables have been measured in the treatment groups in order to make it possible to draw inferences about differences in all parameter estimates for the groups. The error variables, $\epsilon_i^{(g)}$, on the other hand, are indexed, since it is often reasonable to assume that the measured variables have different reliabilities in different groups, and if this fact is not taken into account these reliability differences may cause false conclusions (cf. Campbell, 1963).

If it is assumed that $E(\epsilon_i^{(g)}) = 0$ for all groups it follows from (2) that

$$(3) \qquad E(x_i^{(g)}) = \mu_i + \lambda_i \theta^{(g)} ,$$

where

$$\theta^{(g)} = E(\xi^{(g)}).$$

From (3) it can be seen that all μ- and θ-parameters cannot be identified simultaneously. We can add a constant a_i to μ_i and compensate for this by subtracting a_i/λ_i from $\theta^{(g)}$ for g=1,2,...G. By this operation the observable variables have not been changed. Thus, in the following we have fixed $\theta^{(1)}$ to be equal to zero. When estimates have been obtained we are free to do any rescaling, e.g. such that $\sum_{g=1}^{G} N^{(g)} \theta^{(g)} = 0$, where $N^{(g)}$ is the sample size of group g. For simplicity we can assume that the criterion variable, $\eta^{(g)}$, has been measured without error by a variable $y^{(g)}$. As noted by Lord (1960) this does not introduce any severe restrictions into the model, since an error in the y-variable can be regarded as being absorbed by the error in equation variable, $\zeta^{(g)}$, in (1). Thus, if it is assumed that $E(\zeta^{(g)}) = 0$ for all groups it follows from (1)

$$(4) \qquad E(y^{(g)}) = E(\mu_3 + \eta^{(g)}) = \mu_3 + \alpha^{(g)} + \gamma^{(g)}\,\theta^{(g)} .$$

In (4) we can add a constant to μ_3 and compensate for this by subtracting the same constant from $\alpha^{(g)}$ for all groups. This does not alter the observable variables $y^{(g)}$, so for the same reason as for the $x^{(g)}$-variables we have to fix at least one $\alpha^{(g)}$. In the following $\alpha^{(1)}$ is set to zero.

If it is assumed that the error variables $\epsilon_1^{(g)}$ and $\epsilon_2^{(g)}$ are uncorrelated and each is uncorrelated with $\xi^{(g)}$, the variance-covariance matrix for the observable variables, $x_1^{(g)}$, $x_2^{(g)}$, and $y^{(g)}$, is given by

$$(5) \qquad \underset{\sim}{\Sigma}^{(g)} = \begin{bmatrix} \sigma^2_{\xi^{(g)}} + \sigma^2_{\epsilon_1^{(g)}} & & \\ \lambda_2\sigma^2_{\xi^{(g)}} & \lambda_2^2\sigma^2_{\xi^{(g)}} + \sigma^2_{\epsilon_2^{(g)}} & \\ \gamma^{(g)}\sigma^2_{\xi^{(g)}} & \lambda_2\gamma^{(g)}\sigma^2_{\xi^{(g)}} & {\gamma^{(g)}}^2\sigma^2_{\xi^{(g)}} + \sigma^2_{\zeta^{(g)}} \end{bmatrix}$$

We have three observable variables, which means that for each group there are 9 observable parameters, 6 variances and covariances and 3 means, as long as we restrict our interest to first and second order moments. In the model there are $7G + 2$ parameters, $3G$ parameters from the regression functions, $\sigma^2_{\xi^{(g)}}$, $\gamma^{(g)}$, and $\sigma^2_{\zeta^{(g)}}$, $2G$ parameters for the error variances, $2(G-1)$ parameters for $\theta^{(g)}$ and $\alpha^{(g)}$, since $\theta^{(1)}$ and $\alpha^{(1)}$ are restricted to be equal to zero, 3 μ-parameters and 1 λ-parameter. Thus there are at least $9G - 7G - 2 = 2(G - 1)$ overidentifying restrictions on the model. This means that the estimation of the parameters cannot be done unequivocally by identifying $\underset{\sim}{\Sigma}^{(g)}$ with the sample variance-covariance matrix and the population means with the sample means.

The estimation problem is treated for the general model in section 4.

The model is identified, which can be shown as follows. Let the elements of $\underset{\sim}{\Sigma}^{(g)}$ be denoted by $\sigma_{ij}^{(g)}$, $i,j=1,2,3$. $\gamma^{(g)}$ can be obtained as $\sigma_{32}^{(g)}/\sigma_{21}^{(g)}$ and $\sigma^2_{\xi^{(g)}}$ as $\sigma_{31}^{(g)}/\gamma^{(g)} = \sigma_{31}^{(g)}\sigma_{21}^{(g)}/\sigma_{32}^{(g)}$. For λ_2 there are G possible ways: $\lambda_2 = \sigma_{32}^{(g)}/\sigma_{31}^{(g)}$, $g=1,2,\ldots,G$. This means that there are G - 1 overidentifying restrictions on λ_2. The parameters $\sigma^{2(g)}_{\epsilon_1}$, $\sigma^{2(g)}_{\epsilon_2}$ and $\sigma^{2(g)}_{\zeta}$ can be obtained from $\sigma_{11}^{(g)}$, $\sigma_{22}^{(g)}$ and $\sigma_{33}^{(g)}$ respectively. The means μ_1, μ_2, and μ_3 can be obtained from $E(x_1^{(1)})$, $E(x_2^{(1)})$ and $E(y^{(1)})$ respectively and $\theta^{(g)}$ and $\alpha^{(g)}$ from $E(x_1^{(g)})$ and $E(y^{(g)})$ respectively for $g=2,3,\ldots,G$. Now all parameters have been identified and still we have not used $E(x_2^{(g)})$, $g=2,3,\ldots,G$. Thus, in total there are $2(G-1)$ overidentifying restrictions in the model. There are several possible ways of calculating estimates of the parameters using the observed variances, covariances, and means. In the next section the maximum likelihood method is suggested. With this method all sample information is used, and if the distributional assumptions are met, the estimates are guaranteed to be the most efficient in large samples.

The direct use of the observed variances and covariances is hazardous (cf. e.g. Lord, 1960; Cochran, 1968; Jöreskog and Sörbom, 1974). Suppose we are using y and x_1. Then γ is estimated by

$$\tilde{\gamma}_1 = \text{cov}(x_1, y)/\text{var}(x_1) ,$$

but

$$E(\tilde{\gamma}_1) = \gamma\, \rho_{x_1x_1} ,$$

where

$$\rho_{x_1x_1} = \sigma^2_\xi/(\sigma^2_\xi + \sigma^2_{\epsilon_1}) ,$$

so that $\tilde{\gamma}_1$ is biased downwards. This in turn implies that the estimated treatment effect, $\tilde{\alpha}$, is biased, since

$$\tilde{\alpha} = \bar{y} - \tilde{\gamma}_1\bar{x} .$$

As noted by several authors (see e.g. Smith, 1957; Lord, 1960; Porter, 1971; Bergman, 1972) this bias can have detrimental effects on the conclusions made from an analysis of covariance. Depending on the actual values of $\bar{x}$ in the treatment groups, the bias can lead to a rejection of the hypothesis of no treatment effect when there is no such effect. Also, it can happen that the analysis fails to detect an existing effect because of this bias.

If, for example, we are studying two groups with the same criterion-covariate slope, γ, the effect of the treatment in group 2 as compared with the effect of the treatment in group 1 would have been computed as

$$\tilde{\alpha}^{(2)} - \tilde{\alpha}^{(1)} = \bar{y}^{(2)} - \bar{y}^{(1)} - \tilde{\gamma}(\bar{x}^{(2)} - \bar{x}^{(1)}) .$$

Thus, by (3) and (4)

$$E(\tilde{\alpha}^{(2)} - \tilde{\alpha}^{(1)}) = \alpha^{(2)} - \alpha^{(1)} + \gamma(\theta^{(2)} - \theta^{(1)}) - \gamma\, \rho_{xx}(\theta^{(2)} - \theta^{(1)}) =$$

$$= \alpha^{(2)} - \alpha^{(1)} + \gamma(\theta^{(2)} - \theta^{(1)})(1 - \rho_{xx}) .$$

This means that as long as x is not measured without error, i.e. $\rho_{xx} \neq 1$, we obtain a bias in the estimated effect, and this bias can be positive or negative depending on the difference in level of the covariate among the groups. An obvious way to eliminate the bias is to include an estimate of the reliability of x, r_{xx}, in the analysis. That is, instead of $\tilde{\gamma}$ we use $\tilde{\gamma}/r_{xx}$ (see e.g. Cochran, 1968). In this case the bias equals $\gamma(\theta^{(2)} - \theta^{(1)})(1 - \rho_{xx}/r_{xx})$, and the bias is removed whenever r_{xx} is an exact estimate of ρ_{xx}. However, it is quite seldom that estimates of ρ_{xx} are available, especially when the reliabilities vary among the treatment groups (cf. Campbell, 1963).

3. The general model

In this section we consider a generalization of the model in the previous section. The model is similar to that in Keesling and Wiley (1975) and to the LISREL model (Jöreskog and van Thillo, 1973; Jöreskog and Sörbom, 1976a, 1976b). However, the model to be presented makes it possible to undertake a simultaneous analysis of several groups and to take account of the means of unobserved variables in the model. This is an important feature if we are interested in the effect of treatments. Information about the parameters of the model is contained in the sample variance-covariance matrix as well as in the means of the observed variables, and these two sources of information are not independent. This implies that in order to obtain estimates which are fully efficient one must find them in one step.

Let $\underset{\sim}{y}^{(g)}$ denote a vector of p criterion variables for the g:th treatment group, and $\underset{\sim}{x}^{(g)}$ a vector af q covariates. These covariates may be considered as a set of variables containing information of any kind of preexisting differences among the groups. It is supposed that the $\underset{\sim}{y}^{(g)}$- and $\underset{\sim}{x}^{(g)}$-variables are measuring the unobservable $\underset{\sim}{\eta}^{(g)}$- and $\underset{\sim}{\xi}^{(g)}$-variables according to the following measurement model:

$$
\begin{aligned}
\underset{\sim}{y}^{(g)} &= \underset{\sim}{\mu}_y + \underset{\sim}{\Lambda}_y \underset{\sim}{\eta}^{(g)} + \underset{\sim}{\varepsilon}_y^{(g)} \\
\underset{\sim}{x}^{(g)} &= \underset{\sim}{\mu}_x + \underset{\sim}{\Lambda}_x \underset{\sim}{\xi}^{(g)} + \underset{\sim}{\varepsilon}_x^{(g)}
\end{aligned}
\tag{6}
$$

Our main interest is focused on the parameters of the structural equations for the groups, that is

$$
\underset{\sim}{\eta}^{(g)} = \underset{\sim}{\alpha}^{(g)} + \underset{\sim}{\Gamma}^{(g)} \underset{\sim}{\xi}^{(g)} + \underset{\sim}{\zeta}^{(g)} \tag{7}
$$

Differences of the $\underset{\sim}{\alpha}$-vectors are associated with the usual ANCOVA treatment effects.

Suppose we have $N^{(g)}$ observations of the q-dimensional vector $\underset{\sim}{x}^{(g)}$ and the p-dimensional vector $\underset{\sim}{y}^{(g)}$ for g=1,2,...,G, where G is the number of groups. If it is assumed that the expected means of the error terms $\underset{\sim}{\varepsilon}_{y}^{(g)}$ and $\underset{\sim}{\varepsilon}_{y}^{(g)}$ in (6) are equal to zero it follows that

$$
(8) \qquad \begin{aligned}
E(\underset{\sim}{y}^{(g)}) &= \underset{\sim}{\mu}_{y} + \underset{\sim}{\Lambda}_{y}\, \underset{\sim}{\theta}_{\eta}^{(g)} \\
E(\underset{\sim}{x}^{(g)}) &= \underset{\sim}{\mu}_{x} + \underset{\sim}{\Lambda}_{x}\, \underset{\sim}{\theta}_{\xi}^{(g)} ,
\end{aligned}
$$

where $\underset{\sim}{\theta}_{\eta}^{(g)}$ and $\underset{\sim}{\theta}_{\xi}^{(g)}$ denote the expectations of $\underset{\sim}{\eta}^{(g)}$ and $\underset{\sim}{\xi}^{(g)}$ respectively. If there are no restrictions on $\underset{\sim}{\alpha}^{(g)}$ in (7), this vector is obtained as

$$
(9) \qquad \underset{\sim}{\alpha}^{(g)} = \underset{\sim}{\theta}_{\eta}^{(g)} - \Gamma\, \underset{\sim}{\theta}_{\xi}^{(g)} .
$$

All the $\underset{\sim}{\theta}_{\eta}^{(g)}$- and $\underset{\sim}{\theta}_{\xi}^{(g)}$-vectors cannot be identified, since analogous to the case considered in Section 2, we can add a vector $\underset{\sim}{a}_{\eta}$ to $\underset{\sim}{\theta}_{\eta}^{(g)}$, and a vector $\underset{\sim}{a}_{\xi}$ to $\underset{\sim}{\theta}_{\xi}^{(g)}$ in (8) for g=1,2,...,G and compensate for this by subtracting $\underset{\sim}{\Lambda}_{y}\underset{\sim}{a}_{\eta}$ and $\underset{\sim}{\Lambda}_{x}\underset{\sim}{a}_{\xi}$ from $\underset{\sim}{\mu}_{y}$ and $\underset{\sim}{\mu}_{x}$ respectively. In the following these indeterminacies are circumvented by letting $\underset{\sim}{\theta}_{\eta}^{(1)} = \underset{\sim}{0}$ and $\underset{\sim}{\theta}_{\xi}^{(1)} = \underset{\sim}{0}$. Afterwards, when estimates of the parameters are obtained, we are free to make any translation of the above type. For example, it can be done in such a way that

$$
(10) \qquad \begin{aligned}
\sum_{g=1}^{G} N^{(g)} \underset{\sim}{\alpha}^{(g)} &= 0 \\
\sum_{g=1}^{G} N^{(g)} \underset{\sim}{\theta}_{\xi}^{(g)} &= 0
\end{aligned}
$$

To simplify notation let

$$
\underset{\sim}{z}^{(g)} = \begin{pmatrix} \underset{\sim}{y}^{(g)} \\ \underset{\sim}{x}^{(g)} \end{pmatrix}, \quad \underset{\sim}{\Lambda} = \begin{pmatrix} \underset{\sim}{\Lambda}_{y} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Lambda}_{x} \end{pmatrix}, \quad \underset{\sim}{\omega}^{(g)} = \begin{pmatrix} \underset{\sim}{\eta}^{(g)} \\ \underset{\sim}{\xi}^{(g)} \end{pmatrix}, \quad \underset{\sim}{\varepsilon}^{(g)} = \begin{pmatrix} \underset{\sim}{\varepsilon}_{y}^{(g)} \\ \underset{\sim}{\varepsilon}_{x}^{(g)} \end{pmatrix},
$$

$$\underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu}_y \\ \underset{\sim}{\mu}_x \end{pmatrix}, \quad \underset{\sim}{\theta}^{(g)} = \begin{pmatrix} \underset{\sim}{\theta}_\eta^{(g)} \\ \underset{\sim}{\theta}_\xi^{(g)} \end{pmatrix}.$$

Then (6) can be written as

$$\underset{\sim}{z}^{(g)} = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\,\underset{\sim}{\omega}^{(g)} + \underset{\sim}{\varepsilon}^{(g)}\,, \tag{11}$$

and (8) as

$$E(\underset{\sim}{z}^{(g)}) = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\,\underset{\sim}{\theta}^{(g)}\,. \tag{12}$$

If it is assumed that $\underset{\sim}{\varepsilon}^{(g)}$ is uncorrelated with $\underset{\sim}{\omega}^{(g)}$ and has expectation $\underset{\sim}{0}$, then it follows that the variance-covariance matrix for $\underset{\sim}{z}^{(g)}$ equals

$$\begin{aligned} \underset{\sim}{\Sigma}^{(g)} &= E[\underset{\sim}{z}^{(g)} - E(\underset{\sim}{z}^{(g)})]\,[\underset{\sim}{z}^{(g)} - E(\underset{\sim}{z}^{(g)})]' = \\ &= E(\underset{\sim}{z}^{(g)} - \underset{\sim}{\mu} - \underset{\sim}{\Lambda}\,\underset{\sim}{\theta}^{(g)})(\underset{\sim}{z}^{(g)} - \underset{\sim}{\mu} - \underset{\sim}{\Lambda}\,\underset{\sim}{\theta}^{(g)})' = \\ &= E[\underset{\sim}{\Lambda}(\underset{\sim}{\omega}^{(g)} - \underset{\sim}{\theta}^{(g)}) + \underset{\sim}{\varepsilon}^{(g)}]\,[\underset{\sim}{\Lambda}(\underset{\sim}{\omega}^{(g)} - \underset{\sim}{\theta}^{(g)}) + \underset{\sim}{\varepsilon}^{(g)}]' = \\ &= \underset{\sim}{\Lambda}\,\underset{\sim}{\Phi}^{(g)}\,\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}^{(g)}\,, \end{aligned} \tag{13}$$

where $\underset{\sim}{\Phi}^{(g)}$ and $\underset{\sim}{\Psi}^{(g)}$ are the variance-covariance matrices for $\underset{\sim}{\omega}^{(g)}$ and $\underset{\sim}{\varepsilon}^{(g)}$ respectively. In each group the model for $z^{(g)}$ is similar to a restricted factor analysis model (cf. Lawley and Maxwell, 1971). However, it should be noted that there is no assumption of diagonality for the $\underset{\sim}{\Psi}^{(g)}$-matrices. This means that the model allows for covariances among the $\underset{\sim}{\varepsilon}^{(g)}$-variables in (11), and this feature of the model can be of importance for some special data designs; see Section 5 for an example.

If it is further assumed that $\underset{\sim}{\zeta}^{(g)}$ in (7) is uncorrelated with $\underset{\sim}{\varepsilon}^{(g)}$ and $\underset{\sim}{\xi}^{(g)}$ and has expectation $\underset{\sim}{0}$ it follows that $\underset{\sim}{\Phi}^{(g)}$ has the structure

$$\underset{\sim}{\Phi}^{(g)} = \begin{bmatrix} \underset{\sim}{\Gamma}^{(g)}\,\underset{\sim}{\Phi}_{\xi\xi}^{(g)}\,\underset{\sim}{\Gamma}'^{(g)} + \underset{\sim}{\Theta}^{(g)} & \\ \underset{\sim}{\Phi}_{\xi\xi}^{(g)}\,\underset{\sim}{\Gamma}'^{(g)} & \underset{\sim}{\Phi}_{\xi\xi}^{(g)} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\Phi}_{\eta\eta}^{(g)} & \\ \underset{\sim}{\Phi}_{\xi\eta}^{(g)} & \underset{\sim}{\Phi}_{\xi\xi}^{(g)} \end{bmatrix}, \tag{14}$$

where $\underset{\sim}{\Theta}^{(g)}$ is the variance-covariance matrix for the errors $\underset{\sim}{\zeta}^{(g)}$ in (7).

4. Estimation of the model

As noted above the model is similar to a factor analysis model for several groups, where the means of the factors are taken into account. This model has been treated in Sörbom (1974), and the only difference here is that the variance-covariance matrix for the factors has the structure (14). The estimation of the parameters of the model can be treated in the same fashion as in Sörbom (1974), where the modified Fletcher and Powell (1963) procedure as modified by Gruvaeus and Jöreskog (1970) was used.

Let $\underset{\sim}{z}_i^{(g)}$ denote the i:th observation in the g:th group. Assuming that $\underset{\sim}{z}^{(g)}$ has a multinormal distribution and that the observations are obtained independently, it follows that minus the natural logarithm of the likelihood function for the g:th group is given by

$$(15) \qquad f^{(g)} = (N^{(g)}/2)\,[\log|\underset{\sim}{\Sigma}^{(g)}| + \mathrm{tr}(\underset{\sim}{\Sigma}^{-1^{(g)}}\,\underset{\sim}{T}^{(g)})] ,$$

where $|\cdot|$ denotes the determinant and tr(.) the trace of a matrix. $\underset{\sim}{T}^{(g)}$ in (15) is the matrix

$$(16) \qquad \underset{\sim}{T}^{(g)} = 1/N^{(g)} \sum_{i=1}^{N^{(g)}} (\underset{\sim}{z}_i^{(g)} - \underset{\sim}{\mu} - \underset{\sim}{\Lambda}\,\underset{\sim}{\theta}^{(g)})(\underset{\sim}{z}_i^{(g)} - \underset{\sim}{\mu} - \underset{\sim}{\Lambda}\,\underset{\sim}{\theta}^{(g)})' .$$

The maximum likelihood (ML) estimates of the parameters of the model are defined as those values of the parameters that make the function

$$(17) \qquad F = \sum_{g=1}^{G} f^{(g)}$$

attain its minimum. With the Fletcher and Powell procedure, the minimum of (17) is obtained by an iterative algorithm which make use of the first derivatives. These derivatives are given in Jöreskog (1971a) and Sörbom (1974) except for the parameters in $\underset{\sim}{\Phi}_{\xi\xi}^{(g)}$, $\underset{\sim}{\Gamma}^{(g)}$ and $\underset{\sim}{\Theta}^{(g)}$ in (14). It can be shown in a similar manner as in Jöreskog (1973) that these are given by

$$(1/N^{(g)})\delta F/\delta\ \underset{\sim}{\Phi}_{\xi\xi}^{(g)} = \underset{\sim}{\Gamma}'^{(g)}\ \underset{\sim}{\Lambda}_y'\ \underset{\sim}{\Omega}_{yy}^{(g)}\ \underset{\sim}{\Lambda}_y\ \underset{\sim}{\Gamma}^{(g)} + \underset{\sim}{\Lambda}_x'\ \underset{\sim}{\Omega}_{xy}^{(g)}\ \underset{\sim}{\Lambda}_y \underset{\sim}{\Gamma}^{(g)} +$$

$$+ \underset{\sim}{\Gamma}'^{(g)}\ \underset{\sim}{\Lambda}_y'\ \underset{\sim}{\Omega}_{yx}^{(g)}\ \underset{\sim}{\Lambda}_x + \underset{\sim}{\Lambda}_x'\ \underset{\sim}{\Omega}_{xx}^{(g)}\ \underset{\sim}{\Lambda}_x$$

(18)

$$(1/N^{(g)})\delta F/\delta\ \underset{\sim}{\Gamma}^{(g)} = \underset{\sim}{\Lambda}_y'(\underset{\sim}{\Omega}_{yy}^{(g)}\ \underset{\sim}{\Lambda}_y\ \underset{\sim}{\Gamma}^{(g)} + \underset{\sim}{\Omega}_{yx}^{(g)}\ \underset{\sim}{\Lambda}_x)\ \underset{\sim}{\Phi}_{\xi\xi}^{(g)}$$

$$(1/N^{(g)})\delta F/\delta\ \underset{\sim}{\Theta}^{(g)} = \underset{\sim}{\Lambda}_y'\ \underset{\sim}{\Omega}_{yy}^{(g)}\ \underset{\sim}{\Lambda}_y$$

where $\underset{\sim}{\Omega}^{(g)}$ denotes the matrix

$$\underset{\sim}{\Omega}^{(g)} = \begin{pmatrix} \underset{\sim}{\Omega}_{xx}^{(g)} & \underset{\sim}{\Omega}_{xy}^{(g)} \\ \underset{\sim}{\Omega}_{yx}^{(g)} & \underset{\sim}{\Omega}_{yy}^{(g)} \end{pmatrix} = \underset{\sim}{\Sigma}^{-1(g)}(\underset{\sim}{\Sigma}^{(g)} - \underset{\sim}{T}^{(g)})\underset{\sim}{\Sigma}^{-1(g)}.$$

As in Sörbom (1974) the parameters of the model are divided into three categories

(i) fixed parameters, i.e. parameters specified to have a given value;

(ii) free parameters, i.e. parameters the value of which are unknown and are to be estimated from data;

(iii) constrained parameters, i.e. parameters specified to be equal to one or more other parameters.

By this division of the parameters it is possible to specify a wide variety of different models and to take into account prior information about the data.

The value of F in (17) at the minimum can be used in hypothesis testing. For example, in the case discussed in Section 2, the equality of the slopes in the regressions of criterion variable on true covariate, that is the $\gamma^{(g)}$ in (1), can be tested in the following manner: first estimate the model with no restrictions on the $\gamma^{(g)}$-parameters. This gives a value of F equal to F_0, say. Then we can estimate the model with the restriction $\gamma^{(1)} = \gamma^{(2)} = \dots = \gamma^{(G)}$, which gives rise to another value of F, F_1, say. This value is greater than F_0 and we can examine the difference F_1-F_0 to see whether the hypothesis of equal γ:s is acceptable. In fact, the difference is the likelihood ratio test statistic, and for large samples F_1-F_0 is approximately distributed as χ^2 with G-1 degrees of freedom.

5. An example

To illustrate the use of the model and a procedure for model modification, a small subset of data from a study conducted by Sten Olsson (1973) is analysed in this section. The example is chosen mainly for the purpose of clarifying the use of the model and to illustrate how the parameters of the model can be interpreted in a real situation. In fact, the data consist of the smallest possible set for an analysis of the kind proposed, and, no doubt, the degrees of freedom for the models analysed are too small to make any wider conclusions. Rather, the analysis reported should be regarded as pure illustrations, which for the sake of clarity have been chosen to be as simple as possible.

The main data in the Sten Olsson (1973) study consist of eight tests from the DBA-test battery (Härnqvist, 1962). These tests were administered to about 400 11-year old pupils at two occasions, approximately one month apart. During this timeperiod four groups of about 100 pupils each were given different degree of training in two of the verbal tests, Synonyms (S) and Opposites (O). The children were in principle randomly assigned to the treatment groups. In the example data from two of the groups, Experiment group 1 (E) and Control group (C), were chosen. The members of the E-group were given tests similar to the S- and O-tests three times between the two administrations. These training tests contained all the items in the original test plus a number of similar items. The pupils were given the correct answers of the items after each training session. For group C, there was no such training.

TABLE 1. Sample variance-covariance matrices and sample means for the data used in the example.

Control Group (N=105)

Test	S				Means
$X_S^{(C)}$	37.626				18.381
$X_0^{(C)}$	24.933	34.680			20.229
$Y_S^{(C)}$	26.639	24.236	32.013		20.400
$Y_0^{(C)}$	23.649	27.760	23.565	33.443	21.343

Experiment Group (N=108)

Test	S				Means
$X_S^{(E)}$	50.084				20.556
$X_0^{(E)}$	42.373	49.872			21.241
$Y_S^{(E)}$	40.760	36.094	51.237		25.667
$Y_0^{(E)}$	37.343	40.396	39.890	53.641	25.870

The sample variance-covariance matrices and the sample means for the two groups are given in Table 1.

The tests S and O are both assumed to measure verbal ability, so let us assume that the common part of these tests is a factor which we can call verbal ability (VA). Intuitively the model shown in the path diagram in Figure 1 should be adopted first. In the figure observed variables are enclosed in squares and unobserved variables in circles. An arrow between variables indicates a possible direct causal influence. Thus, for example, the score of the S-test for an individual is supposed to be composed of the individual's verbal ability and an error term, ϵ. The error should here be interpreted in a rather broad sense. It is that part of the test score that remains after the influence of the verbal ability has been removed. Thus, apart from what is usually meant by measurement errors, the ϵ-variable also contains influence from other abilities and traits of the individual which are not involved in other tests measuring the same ability.

In accordance with the general model in (6) we can explicitly write the model for the test scores in the g:th group as

$$(19)\qquad \begin{aligned} x_S^{(g)} &= \mu_1 + \xi^{(g)} + \epsilon_1^{(g)} \\ x_O^{(g)} &= \mu_2 + \lambda_2\xi^{(g)} + \epsilon_2^{(g)} \\ y_S^{(g)} &= \mu_3 + \eta^{(g)} + \epsilon_3^{(g)} \\ y_O^{(g)} &= \mu_4 + \lambda_4\eta^{(g)} + \epsilon_4^{(g)} , \end{aligned}$$

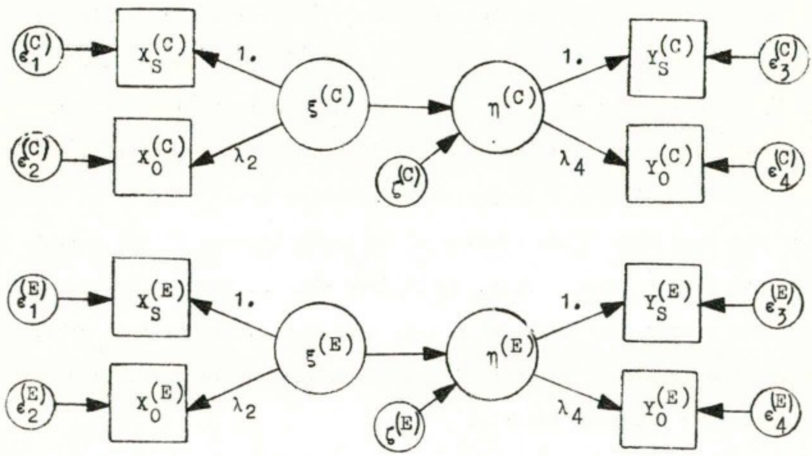

FIGURE 1. The initial model for the two groups in the example.

where $x_S^{(g)}$ and $y_S^{(g)}$ denote test scores for the Synonyms test at the first and second occasion, respectively, and $x_O^{(g)}$ and $y_O^{(g)}$ denote the corresponding scores for the Opposites test. $\xi^{(g)}$ and $\eta^{(g)}$ are the verbal abilities at the first and second occasion, respectively. The superscript, g, has the value E for the Experiment group and C for the Control group. The structural equation to be studied is given by

$$(20) \qquad \eta^{(g)} = \alpha^{(g)} + \gamma^{(g)} \xi^{(g)} + \zeta^{(g)} .$$

When the model in Figure 1 is estimated by the method outlined in Section 4, this results in an overall chi-square measure of fit equal to 35.1 with 6 degrees of freedom. The differences between the observed variance-covariance matrices and the observed means and the estimated Σ-matrices in (13) and estimated expected values in (12) are rather large. Thus, it seems that the model cannot be used to describe the data sufficiently well. The model has to be modified in some way. In Sörbom (1975) a procedure is described which uses the derivatives of the fixed parameters of the function F in (17) to indicate in what respect the model has been specified wrongly. A study of these derivatives suggests that there exists a covariance for the error of the O-test between the two occasions in the E-group. A covariance of this kind can be interpreted in several ways. For example, the O-test may contain a measure of some ability not contained in the S-test, and as the same tests were used at the two occasions, it seems natural that there is some remaining correlation between the test scores for the O-test after the influence of the true scores has been removed (cf. Sörbom, 1975). Thus, when this covariance between the errors has been taken into consideration by the model we assume that the verbal ability factor should have been more "pure". The model with the covariance included yields a chi-square with 5 degrees of freedom equal to 17.2. The decrease in chi-square for this model as compared with the initial model equals 17.9 (=35.1-17.2). Thus, the hypothesis of zero covariance between the errors for the O-test in the E-group is rejected by a chi-square with 1 degree of freedom equal to 17.9. Still, the overall fit of the model is not acceptable in a strict interpretation of the chi-square measure, and an inspection of the above-mentioned derivatives for the model shows that there might be a covariance of the O-test errors also in the C-group. Allowing also this covariance to be a free parameter we obtain a model with an overall fit measure equal to 2.8. As for the E-group we can conclude that there is a covariance for errors in the C-group. This time we can reject the hypothesis of no covariance by virtue of a chi-square with 1 degree of freedom equal to 14.4 (=17.2-2.8). The overall fit of the model is now very good, the probability level for a chi-square equal to 2.8 with 4 degrees of freedom is approximately 0.59, and the sample variance-covariance matrix and the sample means are reproduced by the model parameter estimates to at least two significant digits.

To make inferences about differences among the groups in the development of verbal ability as measured by the two tests, we can use the model with $\sigma_{\epsilon_2\epsilon_4}$ free for the two groups. Our main interest is to compare the estimated structural equations (20). An inspection of these reveals that the γ-parameter, i.e. the slope in the regression of true criterion score on true covariate score, are almost equal, 0.947 for the C-group and 0.854 for the E-group. The hypothesis that these slopes are equal can be tested by estimating the model once more but with the restriction $\gamma^{(C)} = \gamma^{(E)}$ added. This results in a model with an overall chi-square measure equal to 4.0 with 5 degrees of freedom. Thus, the hypothesis of equal slopes cannot be rejected, since it leads to a chi-square with 1 degree of freedom equal to 1.2, and this is not significant. The different models and their associated chi-square values are summarized in Table 2. The maximum likelihood estimates of the parameters in the final model, with equal criterion-covariate slopes, are listed in Table 3.

TABLE 2. Measures of overall fit for the models analysed in the example.

Model	chi-square	d.f.	Probability level
Initial	35.056	6	0.000
$\sigma^{(E)}_{\epsilon_2\epsilon_4}$ free	17.187	5	0.004
$\sigma^{(E)}_{\epsilon_2\epsilon_4}$ and $\sigma^{(C)}_{\epsilon_2\epsilon_4}$ free	2.799	4	0.592
$\sigma^{(E)}_{\epsilon_2\epsilon_4}$ and $\sigma^{(C)}_{\epsilon_2\epsilon_4}$ free, $\gamma^{(E)}=\gamma^{(C)}$	3.989	5	0.551

The above results can be compared with a more traditional analysis of the data. First of all we want to see whether there is a difference among the groups at the post-test occasion. To do this we can conduct a usual t-test, that is using the test statistic

$$(21) \qquad t = (\bar{y}^{(C)} - \bar{y}^{(E)})/(s_{yy}^{(C)}/N^{(C)} + s_{yy}^{(E)}/N^{(E)})^{1/2} ,$$

where $\bar{y}^{(g)}$ denotes the sample mean for either the S- or O-test and $s_{yy}^{(g)}$ the sample variance. Because the sample sizes are fairly large, t is in the following regarded as being normally distributed. For the data in Table 1 the test results in a t equal to -5.97 for the S-test and -5.01 for the O-test. However, in the test we have overestimated the variances in the denominator of (21), since the test score variances also contains measurement errors. Thus, a better test should be to use the true post-test score differences. By use of the estimates in Table 3 we obtain a t-value equal to $-5.306/(24.897/105 + 46.738/108)^{1/2} = -6.482$. Thus, this test gives a stronger support for the rejection of the hypothesis of no group differences at the post-test occasion.

However, as can be seen from Table 1, the level of the E-group test means were higher at the pre-test occasion, too. In fact, a t-test of the difference in mean values for the S-test among the groups results in a t-value equal to -2.51, which is significant at least at the 1 per cent level. For the O-test the t-value is -1.19 and this is not significant. On the other hand, if we use the results from the maximum likelihood estimation in Table 3, to test for difference in true score mean at the pre-test occasion, we obtain a t-value equal to -3.59, which is highly significant. In this case we have gained in power for the same reason as in the analysis of the post-test occasion.

TABLE 3. Maximum likelihood estimates of the parameters for the final model in the example.

(Fixed parameters are denoted by an asterisk, *)

Parameter	Control Group		Experiment Group
λ_2		0.878	
λ_4		0.907	
σ^2_ξ	29.794		47.334
γ		0.895	
σ^2_ζ	1.032		8.823
$\sigma^2_{\epsilon_1}$	9.584		2.547
$\sigma^2_{\epsilon_2}$	12.030		12.359
$\sigma^2_{\epsilon_3}$	5.836		7.451
$\sigma^2_{\epsilon_4}$	12.500		17.209
$\sigma_{\epsilon_2\epsilon_4}$	6.391		7.304
$E(\xi)$	0.000*		1.875
$E(\eta)$	0.000*		5.306
μ_1		18.619	
μ_2		19.910	
μ_3		20.383	
μ_4		21.203	

Thus, there could have been preexisting differences among the groups which might explain the outcome of the experiment. By the use of the analysis of covariance strategy we can even out these differences. Instead of testing for differences in post-test means we test for differences in post-test means after elimination of the linear regression on pre-test scores by a t-test. For the data in the example this results in a t-value equal to -6.50 and -6.70 for the S- and O-test, respectively. Thus, we have gained in power as compared with the t-test without taking pre-test differences into account. A still more powerful comparison would be obtained if we use the true score estimates in Table 3 to perform an analysis of covariance. The value of the t statistic equals -11.67 in this case and, thus, there is a considerable gain in power as compared with the initial test for differences in the post-test means.

6. Lord's numerical example

In Lord (1960), a large sample test of treatment effects for the case of a fallible covariate is derived. In an example, it was shown that by accounting for the measurement errors in the covariate a significant treatment effect was demonstrated, whereas an ordinary analysis of covariance would not have detected such an effect. In this section the example will be re-analysed to illustrate the specification of the general model in that case, and how a test similar to Lord's is derived.

The data for the example are given in Table 4. They have been taken from a study by Frederiksen and Schrader (1951). The criterion variable, y, is Freshman average grade and there are two covariates, x_1 and x_2. x_2 is not actually observed, since no duplicate covariate was available. It is computed to be a measure parallel to x_1. There were two groups,

TABLE 4. Sample variance-covariance matrices and sample means for the data in Lord's numerical example.

Male Group (N=119)

Variable	S			Means
$x_1^{(1)}$	5.29000			4.07000
$x_2^{(1)}$	4.23200	5.29000		4.07000
$y^{(1)}$	1.12120	1.12120	0.56250	1.40000

Female Group (N=93)

Variable	S			Means
$x_1^{(2)}$	3.88090			5.34000
$x_2^{(2)}$	2.82290	3.88090		5.34000
$y^{(2)}$	0.55278	0.55278	0.37210	1.57000

Male and Female, with 119 and 93 observations, respectively, and x_2 was constructed in such a way that the reliability of x_1 and x_2 was 0.8 in the Male group and 0.72738 in the Female group.

The model for the data is the same as that one discussed in Section 2, except that, because of the construction of the numerical data, $\lambda_1 = \lambda_2 = 1$ in Equation (2) and $\sigma^2_{\epsilon_1}(g) = \sigma^2_{\epsilon_2}(g) = \sigma_{\epsilon}(g)$, say, for $g=1,2$. It is then seen from (5) that we can estimate σ^2_{ξ} by $s_{x_1x_2}$, γ by $s_{yx_1}/s_{x_1x_2}$, σ^2_{ϵ} by $s_{x_1x_1} - s_{x_1x_2}$, and σ^2_{ζ} by $s_{yy} - s^2_{yx_1}/s_{x_1x_2}$ in both groups to obtain estimates that perfectly fit the data. The estimates of γ are quite similar in the two groups; it equals $1.1212/4.232 = 0.2649$ in the Male group and $0.55278/2.8229 = 0.1958$ in the Female group. To test whether the γ:s are equal we estimate the model with the constraint $\gamma^{(1)} = \gamma^{(2)}$ as outlined in Section 4. This gives as a result a model with a chi-square with 1 degree of freedom equal to 2.304, which means that we cannot reject the hypothesis at the 10 per cent significance level. The maximum likelihood estimates of the parameters are listed in Table 5. The next step in the procedure is to test whether there is a difference among the groups in Freshman average grade when the differences in the covariate have been taken into account, that is to test whether α in Table 5 is equal to zero. This can be done by estimating the model with the constraint $\alpha^{(2)} = 0$, or equivalently $E(\eta^{(2)}) = \gamma\, E(\xi^{(2)})$, added. By comparing the chi-square for this model with the chi-square for the previous one we obtain a chi-square test of the hypothesis with 1 degree of freedom. Instead of the function F in (17) we will minimize the function

$$H = F + \tau[E(\eta^{(2)}) - \gamma\, E(\xi^{(2)})] \,, \tag{22}$$

TABLE 5. Maximum likelihood estimates of the parameters in the model for Lord's numerical example.

(Fixed parameters are denoted by an asterisk, *)

Parameter	Male Group		Female Group
λ_1		1.000*	
λ_2		1.000*	
σ_ξ^2	4.318		2.691
γ		0.242	
σ_ζ^2	0.273		0.259
σ_ϵ^2	1.041		1.096
α	0.000*		-0.137
μ_1		4.070	
μ_2		4.070	
μ_3		1.400	

where τ denotes a Lagrangian multiplier. The estimation of the model results in a chi-square with 2 degrees of freedom equal to 5.221. Thus, at the 10 per cent level we can reject the hypothesis of no effect by a chi-square with 1 degree of freedom equal to $5.221 - 2.304 = 2.917$. This value of the test statistic should be compared with the corresponding t-value that is obtained from an ordinary analysis of covariance. As reported by Lord (1960) this equals -0.855, corresponding to a chi-square equal to 0.731. Lord's test gives a t-value equal to -1.69 ($t^2 = 2.856$).

7. Summary and conclusion

There is only a small gain in power by the method proposed as compared with Lord's test for the data in the last example. However, the example illustrates in what ways the method is more general than the procedure proposed by Lord and/or the usual analysis of covariance.

To sum up

1. There is no restriction on the number of treatment groups that can be involved. Nor is there any restriction on the number of covariates or the number of criterion variables.
2. The fallible covariates can be parallel, tau-equivalent, or congeneric measures (see Jöreskog, 1971b), or can conform to a factor analysis model.
3. The method can handle fallible criterion variables as well as fallible covariates.
4. There are no requirements of equal variances for criterion variables for given covariates.
5. There is a provision for a wide variety of different tests regarding the parameters of the model.

References

Bergman, L.R. (1972), Change as the dependent variable. Report from the Psychological Laboratories, Supplement 14, The University of Stockholm.

Bock, R.D. (1975), Multivariate statistical methods in behavioral research. New York: McGraw-Hill.

Campbell, D.T. (1963), From description to experimentation: Interpreting trends as quasi-experiments. In Chester W. Harris (Ed.), Problems in measuring change. Madison: The University of Wisconsin Press, 212-242.

Cochran, W.G. (1968), Errors of measurement in statistics. Technometrics, 10, 637-666.

Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972), The dependability of behavioral measurements: Theory of generalizability for scores and profiles: New York: Wiley.

Cronbach, L.J., Rogosa, D., Price, G. & Floden, R. (1976), Analysis of covariance, temptress and deluder? Or angel of salvation? Mimeograph, Stanford Evaluation Consortium, Stanford University.

Elashoff, J.D. (1969), Analysis of covariance: A delicate instrument. American Educational Research Journal, 6, 383-402.

Fletcher, R. & Powell, M.J.D. (1963), A rapidly convergent descent method for minimization. Computer Journal, 6, 163-168.

Frederiksen, N. & Schrader, W.B. (1951), Adjustment to College. Princeton, New Jersey: Educational Testing Service.

Gruvaeus, G.T. & Jöreskog, K.G. (1970), A computer program for minimizing a function of several variables. Research Bulletin 70-14. Princeton, New Jersey: Educational Testing Service.

Härnqvist, K. (1962), Manual till DBA (Manual for DBA). Stockholm: Skandinaviska Testförlaget. (In Swedish)

Jöreskog, K.G. (1971a), Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Jöreskog, K.G. (1971b), Statistical analysis of sets of congeneric tests. Psychometrika, 36, 109-133.

Jöreskog, K.G. (1973), A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 85-112.

Jöreskog, K.G. & Sörbom, D. (1974), Some regression estimates useful in the measurement of change. Research Report 74-2. University of Uppsala, Department of Statistics.

Jöreskog, K.G. & Sörbom, D. (1976a), Statistical models and methods for test-retest situations. In D.N.M. de Gruijter, L.J. Th. van der Kamp & H.F. Crombag (Eds.), Advances in Psychological and Educational Measurement. London: Wiley. In press.

Jöreskog, K.G. & Sörbom, D. (1976b), Statistical models and methods for analysis of longitudinal data. In D.J. Aigner & A.S. Goldberger (Eds.), Latent variables in socioeconomic models. Amsterdam: North Holland Publishing Company. In press.

Jöreskog, K.G. & van Thillo, M. (1973), LISREL - A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. Research Report 73-5, University of Uppsala, Department of Statistics.

Keesling, J.W. & Wiley, D.E. (1975), Measurement error and the analysis of quasi-experimental data, Version III. Mimeograph, University of Chicago.

Lawley, D.N. & Maxwell, A.E. (1971), Factor analysis as a statistical method. 2nd ed. London: Butterworth & Company.

Lord, F.M. (1960), Large sample covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 55, 307-321.

Lord, F.M. (1963), Elementary models for measuring change. In Chester W. Harris (Ed.), Problems in measuring change. Madison: The University of Wisconsin Press, 21-38.

Lord, F.M. (1967), A paradox in the interpretation of group comparisons. Psychological Bulletin, 68, 304-305.

Lord, F.M. & Novick, M.E. (1968), Statistical theories of mental test scores. Reading: Addison-Wesley Publishing Company.

Olsson, S. (1973), En experimentell studie av övningseffekter på testpoäng och faktorstruktur (An experimental study of the effects of training on test scores and factor structure). University of Uppsala, Department of Education. (In Swedish)

Porter, A.C. (1971), How errors of measurement affect ANOVA, regression analysis, ANCOVA and factor analysis. Paper presented at the 1971 AERA convention, New York.

Smith, H.F. (1957), Interpretation of adjusted treatment means and regressions in analysis of covariance. Biometrics, 13, 282-308.

Sörbom, D. (1974), A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

Sörbom, D. (1975), Detection of correlated errors in longitudinal data. British Journal of Mathematical and Statistical Psychology, 28, 138-151.

Sörbom, D. (1976), A statistical model for the measurement of change in true scores. In D.N.M. Gruijter, L.J. Th. van der Kamp & H.F. Crombag (Eds.), Advances in Psychological and Educational Measurement. London : Wiley. In press.