Developments in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 19, Ljubljana: FDV, 2003

Multivariate Modelling with Latent Variables in Experimental Designs with an Application to Forestry Data

Dario Cziráky, Tugomir Filipan, and Anamarija Pisarović[1]

Abstract

The classical experimental design approach to testing for treatment effects in forestry research applies ANOVA and dummy-variable regression techniques to the individual tree development indicators such as total tree height or trunk diameter. An alternative to standard techniques is provided by a more general framework of structural equation modelling, which encompasses most of the classical techniques and additionally allows estimation of a richer class of models. This includes latent variable models that enable simultaneous incorporation of multiple tree development indicators into a single model treating tree development as a latent variable imperfectly measured by the observable tree measures. A problem in classical experimental design is that we can either test for the treatment effects on separate tree development indicators (e.g., height or diameter), or make an attempt to combine multiple indicators into a single (latent) variable and then test for treatment effects on the composite tree development variable. In this paper we apply structural equations methods to experimental forestry data comparing several approaches to treatment effect testing.

1 Introduction

In the forestry-research experimental designs for testing the treatment effects on tree development ANOVA and dummy-variable regression techniques are routinely applied to individual tree development indicators such as total tree height or trunk diameter. Multivariate techniques that can combine several tree development indicators and test treatment effects on all indicators simultaneously are, however, rarely used in the forestry research.

[1] Department for Environmental Sciences, IMO, P.O. Box 303, 10000 Zagreb, Croatia.

On the other hand, structural and functional relationships among different tree development indicators such as diameter and height are extensively researched in the literature (e.g., Arney, 1985, Wensel *et al*., 1987). For recent literature on general forest growth modelling see *inter alia* Botkin (1993), Bossel and Kriger (1994), Vanclay (1994) and Adler (1995). This line of research shows clear relationships among different tree development indicators suggesting that they all measure the same underlying (latent) variable, which is *tree development*. The problem of how to test for treatment effects within classical experimental design immediately follows. Namely, we can either test for the treatment effects on separate tree development indicators (e.g., height or diameter), or make an attempt to combine multiple indicators into a single (latent) variable and then test for treatment effects on the composite tree development variable.

Testing for treatment effects on different indicators separately is most common in practice but problematic in several ways. If multiple indicators are used to separately measure the effects of applied treatments on overall tree development, there is a good chance that the results will be ambiguous. The second approach, based on multivariate modelling, while substantively appealing, is often difficult to apply in practice partly due to the lack of appropriately taken tree measures for such purpose or due to inadequately applied statistical techniques. The methods most commonly used in multivariate modelling with latent variables are based on linear structural equation modelling (SEM), which belong to the general class of covariance structure analysis. Rather powerful methods are available for modelling multivariate Gaussian variables but these methods generally fail to account for complex nonlinearities in the relationships among particular tree development indicators that are themselves not normally distributed. General SEM models can be specialised to most classical experimental techniques such as ANOVA, ANCOVA or MANOVA (Bagozzi, 1977; Bagozzi and Yi, 1989; Kano, 2001) but in addition they account for measurement errors and latent variables. Kano (2001) gives a general expression for the SEM model that encompasses many of the classical experimental techniques.

This paper analyses data containing three indicators of tree development measured in two time points (end of years 1999 and 2001) from an ongoing multi-annual forestry experiment carried out since 1992 at a site in Fuegenberg, Zillertal (Tyrol, Austria). The dependence of multivariate techniques on distributional assumptions is accounted for by detailed preliminary descriptive analysis and data transformations aiming at assuring Gaussian distribution of the modelled variables. A panel (longitudinal) latent variable model is then developed and estimated using maximum likelihood method for the joint 1999 and 2001 sample. After describing the latent scores technique and computing scores for the overall tree development, treatment effects on the overall tree development are tested using dummy variable regression methods. Finally, a multigroup estimation is used to compare model specification and parameter estimates across different groups

where each group corresponds to a subsample of trees grown on parcels subjected to particular treatments (including control).

2 The data

The data we model in this paper comes from an ongoing multi-annual forestry experiment carried out since 1992 at a site in Fuegenberg, Zillertal (Tyrol, Austria) at the altitude of 1450m and thus the natural growth potential of this site is thus expected to be reduced (see Dittmar *et al.*, 2002). The experiment was set in a randomised block design with six variants and four replications thus making a total of 24 trial plots with Spruce trees (*Picea abies* K.). Five soil correctors (see Table 1) were applied (together with the no-treatment control parcel) in each row, thus there were 4 parcels for each treatment. The data consists of three tree development indicators measured in 1999 and 2001: total tree height at the end of the year, annual height increment (growth), and trunk diameter taken at breast height at the end of the year. The shorthand symbols used for the indicators are explained in Table 1.

Table 1: Definitions of variables and treatments.

Tree development indicators
x_1 = total tree height at the end of 1999
x_2 = annual height increment (growth) in 1999
x_3 = trunk diameter taken at breast height at the end of 1999
x_4 = total tree height et the end of 2001
x_5 = annual height increment (growth) in 2001
x_6 = trunk diameter taken at breast height at the end of 2001
Treatments
T_1 = AV-SPS1: Agrarvital with SPS (1800 kg/Ha)
T_2 = AV-SPS2: Agrarvital with SPS (3600 kg/Ha)
T_3 = Biosol (2400 kg/Ha)
T_4 = AV+Biosol: combination of Agrarvital with SPS (3600 kg/Ha) and Biosol (2400 kg/Ha)
T_5 = Biomag (3600 kg/Ha)

From the beginning of the experiment, each individual tree was numbered and thus track was kept of each tree across the years. The number of live trees, however, from year to year was decreasing thus we use the data on trees that were still alive in the end of 2001, the total number of which is 1931. This allowed us to pool the data and form a panel for the two years with matching data for each tree in the sample.

Tests for normality of the distribution (Table 2) reject the null for each of the three variables in both years (for details on these tests see D'Agustino, 1986; Bowman and Shenton, 1975; Doornik and Hansen, 1994; Shenton and Bowman, 1977; Mardia 1980).

Table 2: Normality tests and descriptive statistics (N = 1931).

	1999 data					
	x_1	x_1^{NS}	x_2	x_2^{NS}	x_3	x_3^{NS}
Mean ($\bar{x}$)	102.23	102.23	16.02	16.03	2.85	2.86
Std. dev. (σ)	36.36	36.36	11.69	11.69	0.90	0.90
Skewness	0.58	0.00	0.84	0.01	0.25	0.00
Excess kurtosis	0.04	0.00	0.03	−0.03	−0.23	−0.01
Minimum	31.50	−11.40	0.00	−22.38	1.50	−0.29
Maximum	267.00	233.86	6.00	58.01	6.60	6.00
Normality $\chi^2(2)$	167.89	0.01	46.46	0.02	35.99	0.00
p-value	0.00	0.99	0.00	0.98	0.00	0.73
	2001 data					
	x_4	x_4^{NS}	x_5	x_5^{NS}	x_6	x_6^{NS}
Mean ($\bar{x}$)	126.73	126.73	13.34	13.34	3.54	3.54
Std. Dev. (σ)	55.05	55.05	13.48	13.48	1.20	1.20
Skewness	0.82	0.00	1.56	0.01	0.42	0.00
Excess kurtosis	0.45	−0.01	2.13	−0.04	0.02	−0.01
Minimum	50.00	−24.56	0.00	−28.83	0.90	−0.82
Maximum	324.00	318.02	75.50	6.64	9.90	7.91
Normality $\chi^2(2)$	334.49	0.01	1786.90	0.13	73.46	0.00
p-value	0.00	0.88	0.00	0.58	0.00	0.99

The growth increase is particularly skewed in 1999, with large excess kurtosis (normality χ^2 is 682.3). In 2001 growth was also the most problematically distributed variable with the normality χ^2 statistic of 1262.5. Related to non-normality and an even more serious problem for multivariate techniques we apply in this paper is a notable non-linearity in the relationships (see upper part of Figure 1). Figure 1 includes interpolated non-linear regression curves that signify likely non-linearity in bivariate relationships between pairs of tree measures.

Because of the sensitivity to departures from normality of the maximum likelihood-based multivariate methods that we use in the analysis, and likely non-linearity in the relationships, transformation deserves particular attention and the analysed variables should be approximately normally distributed. In order to normalise the original variables we apply the *normal scores* technique (Jöreskog, 1999, 2000; Jöreskog *et al.*, 2000) which, when applicable, often shows superior

results to standard Box-Cox type of transforms (for technical details on this technique see Cziráky *et al*. 2002a,b; 2003).

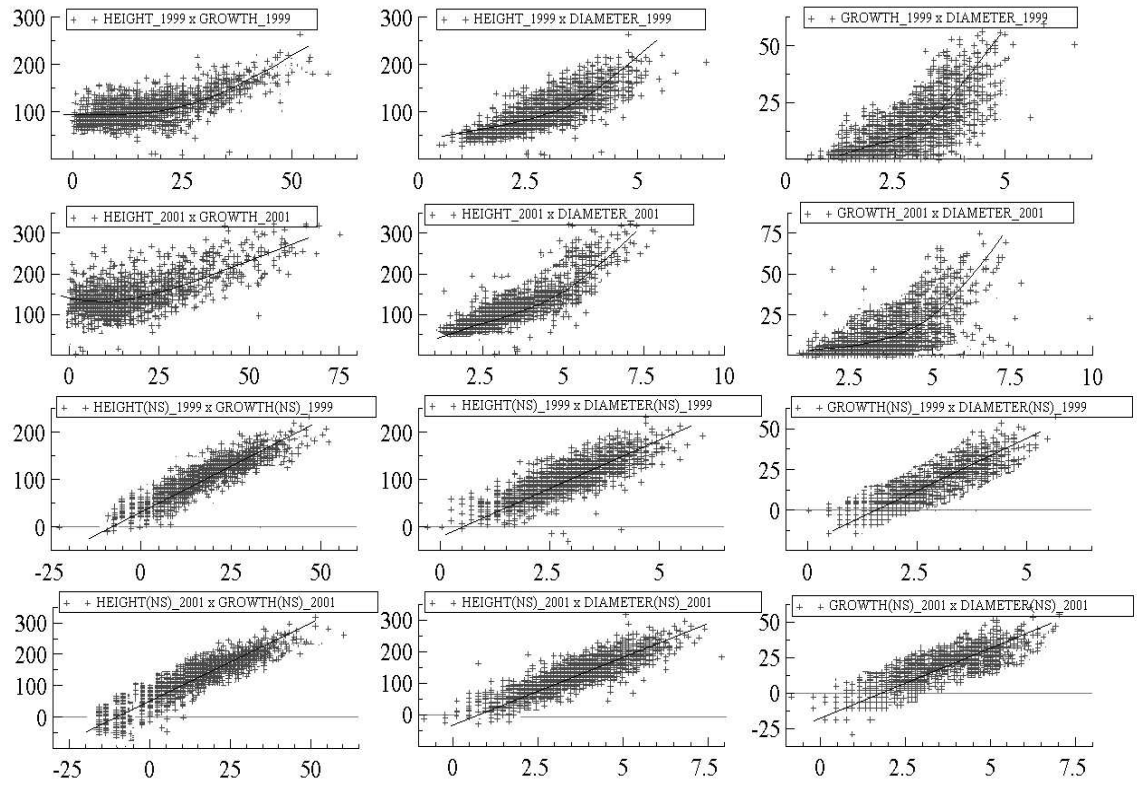

Figure 1: Bivariate cross plots.

In Table 2 variables superscripted "NS" denote normalised variables. It can be observed that none of the χ^2 statistics for the normalised variables is significant, thus the null of Gaussianity cannot be rejected. Skewness and excess kurtosis were significantly reduced by the transformation, noting that the first two empirical moments remained unchanged. Bottom part of Figure 1 shows bivariate plots of transformed variables with interpolated linear regression lines. It is easy to see that normality transformations induced linearity in the relationships, thus in this case the transformations aimed at normalising the data also linearised the relationships. In the following analysis we use normalised variables, thus for simplicity the "NS" superscript will be omitted.

3 Statistical methodology

3.1 Latent variables model

We aim to combine multiple tree development indicators into one single variable that would measure the unobserved overall tree development. The most common approach in multivariate statistics is to estimate a factor-analytic measurement model for tree development where the measures are tree development indicators.

We first define the following notation. Latent variables indicating overall growth progress of the trees for 1999 and 2001 are denoted $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, respectively. The postulated model is an autocorrelated panel model. It is assumed that each of the three tree development indicators are determined by a latent variable denoting "growth progress" of trees, where each indicator is measuring the underlying latent variable with a measurement error δ_i. The panel specification allows the errors of the same indicator across two time periods to correlate (with the covariance parameters θ_{ij}). Note that we do not imply causal dynamics in the latent variables in form of autoregressive (lagged) relationships.

This model can be specified as a special case of the general linear structural equation model with latent variables in the form

$$\mathbf{x} = \boldsymbol{\Lambda}_x \xi + \boldsymbol{\delta} \qquad (1)$$

where $\boldsymbol{\Lambda}_x$ is the matrix of factor loadings and $\mathbf{x}$ and $\boldsymbol{\xi}$ are vectors of observed and latent variables, respectively (see Jöreskog 1973; Jöreskog *et al.,* 2000; Bollen, 1989, Kaplan, 2000). Presence of the residual vector $\boldsymbol{\delta}$ allows for the measurement error in the observed indicators. In matrix notation, the model can be written as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}. \qquad (2)$$

The covariance matrix of the latent variables is given (by writing only the lower triangular elements) as

$$\boldsymbol{\Phi} = E(\xi\xi^T) = \begin{bmatrix} 1 & \\ \phi_{21} & 1 \end{bmatrix} \qquad (3)$$

and $\boldsymbol{\Theta}_\delta$ is the covariance matrix of the residuals. Note that Eq. (3) assumes that latent variables are standardised; reparametrisation of the model can allow for non-standardised $\boldsymbol{\Phi}$ with some loadings in Eq. (2) fixed (usually to unity). The later parametrisation is used to compute scores of the latent variables which then assume the metric of the observed variable with unit loading coefficient. Assuming multivariate Gaussianity, the model parameters can be estimated using *full information maximum likelihood* (FIML) technique (for details see Kaplan, 2000).

3.2 Testing for treatment effects

3.2.1 Latent scores

By computing scores from an estimated latent variables model we get a composite variable that can be used in subsequent analysis specifically in testing for treatment effects on tree development. Of particular interest in this application are methods for estimation of latent scores in the general structural equation models (Jöreskog, 2000). Such methods also allow structural recursive and simultaneous relationships among latent variables. Estimation of factor scores in the pure measurement (factor) models is just a special case of the general procedure (see Lawley and Maxwell, 1971).

We describe a technique capable of computing scores of the latent variables based on the maximum likelihood solution of the autocorrelated panel model following the approach of Jöreskog (2000).

Given the measurement model $\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}$, the latent scores $\boldsymbol{\xi}_i$ can be computed for each observation x_{ij} in the $(6 \times N)$ sample matrix whose rows are observations on each of our six observed variables, i.e.,

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ x_{31} & x_{32} & \cdots & x_{3N} \\ x_{41} & x_{42} & \cdots & x_{4N} \\ x_{51} & x_{52} & \cdots & x_{5N} \\ x_{61} & x_{62} & \cdots & x_{6N} \end{pmatrix} = (\boldsymbol{x}_1\, \boldsymbol{x}_2\, \cdots \boldsymbol{x}_N) \qquad (4)$$

Once the coefficients of $\boldsymbol{\Lambda}_x$ are estimated they can be treated as fixed and the latent scores can be computed by maximising $\sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i)^T \boldsymbol{\Theta}_\delta^{-1} (\mathbf{x}_i - \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i)$ subject to constraint $(1/N) \sum_{i=1}^{N} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T = \boldsymbol{\Phi}$. The solution to this constrained maximisation problem is given by $\boldsymbol{\xi}_i = \mathbf{D}^{1/2} \mathbf{U}^T \mathbf{H} \mathbf{K}^{-1/2} \mathbf{H}^T \mathbf{U} \mathbf{D}^{1/2} \boldsymbol{\Lambda}_x^T \boldsymbol{\Theta}_\delta^{-1} \mathbf{x}_i$, where

$\mathbf{UDU}^T$ and $\mathbf{HKH}^T$ are singular value decompositions of $\mathbf{\Phi}$ and $\mathbf{D}^{1/2}\mathbf{U}^T\mathbf{\Lambda}_x{}^T\mathbf{\Theta}_\delta^{-1}\sum_{i=1}^{N} x_i x_i^T\ \mathbf{\Theta}_\delta^{-1}\mathbf{\Lambda}_x\mathbf{U}\mathbf{D}^{1/2}$, respectively.

Once the latent scores are computed, it is possible to use them in standard (experimental) techniques such as ANOVA or dummy variable regression (see *inter alia* Winer *et al*., 1991). The later technique estimates the treatment effects using ordinary least squares (OLS) and the coefficient vector is given by

$$\boldsymbol{\beta}_{OLS} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\xi}_{\mathrm{i}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}^{1/2}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{K}^{-1/2}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{D}^{1/2}(\boldsymbol{\Lambda}_{\mathrm{x}}{}^{\mathrm{T}}\boldsymbol{\Theta}_{\delta}^{-1}\boldsymbol{x}_{\mathrm{i}}) \qquad (5)$$

where $\mathbf{X}$ is the matrix of dummy variables (including column of 1's for the constant term) indicating treatments. In further analysis we replace control dummy with the constant.

The main advantage of the above approach is that after the latent scores for $\boldsymbol{\xi}_i$'s are estimated they can be used in testing for treatment effects in standard experimental techniques including multiple comparison procedures (see Hochbert and Tamhane, 1987). However, while latent scores in principle offer a convenient summary effect that takes into account multiple tree measures, weighted by their variances and covariances, the standard adjustments to the significance level in multiple comparison procedures (e.g., Šidak, 1967) might have to be further adjusted. The issue of simultaneous testing for differences among treatments (Miller, 1981) would now have to include provision for the fact that the tested latent-scores are obtained on the basis of a multivariate statistical model that might have been build or modified already on inferential grounds. A known problem with the factor scores approach is that, while the estimates of the factor loadings are not biased, the replacement of unknown structural parameters by their estimates induces sample dependence in the estimated factor scores which can affect the accuracy of standard errors and test statistics (Kano, 2001). This indicates an important direction and need for further research, especially in the direction of adjustments of standard errors and fit statistics.

3.2.2 Multiple group estimation

The use of latent scores obtained from multivariate measurement models in ANOVA or regression techniques enables testing for mean differences among treatments. The measurement structure (factor loadings, error covariances and residual covariances) is estimated jointly across the entire sample and is assumed to be equal for all treatments (groups). Alternatively, we can define separate groups based on particular treatment subsamples (i.e., observations coming from parcels treated with particular soil correctors) and estimate model specified in Eq. (2) in each subsample, separately.

Jöreskog (1971) proposed a testing procedure for evaluating group differences in respect to group covariance matrices and group-specific model estimates.

Sörbom (1974) appended Jöreskog's procedure with estimates of latent means which, in principle, allows for estimation of general differences across groups or treatments (see also Sörbom, 1981; Bollen, 1989; Kaplan, 2000). Specifically for our application, we can test three sets of hypotheses, either jointly or separately: $\mathbf{\Lambda}_x^{(1)} = \mathbf{\Lambda}_x^{(2)} = \cdots = \mathbf{\Lambda}_x^{(6)}$, $\mathbf{\Phi}^{(1)} = \mathbf{\Phi}^{(2)} = \cdots = \mathbf{\Phi}^{(6)}$ and $\mathbf{\Theta}_\delta^{(1)} = \mathbf{\Theta}_\delta^{(2)} = \cdots = \mathbf{\Theta}_\delta^{(6)}$, where the numbers in the superscript denote treatment group. Note that the control group is also included as number one treatment (thus there are 6 groups). Formally, the testing proceeds by specifying a multigroup version of the model in Eq. (1) with group (treatment) specific subscript *i*, i.e.,

$$\mathbf{x}_i = \mathbf{\Lambda}_{xi}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i \qquad (6)$$

The hypothesis of overall equality of covariance matrices $\mathrm{H}_0^{(1)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = ... = \mathbf{\Sigma}_k$ can be tested by the Box-*M* test (see Kaplan, 2000). In respect to specific multivariate structure of the analysed matrices, if they are found not to be overall identical we can test for the equality of the number of factors and equality of the model parameters. Particularly, as already mentioned, we can test for $\mathrm{H}_0^{(2)} : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = ... = \mathbf{\Lambda}_k$ by minimising the multi-group likelihood function with and without the equality constraints which allows computing a likelihood ratio statistic and formal hypothesis testing.

The addition of latent means results with inclusion of additional parameters (see Sörbom, 1974) and the model becomes $\mathbf{x}_i = \boldsymbol{\tau}_i + \mathbf{\Lambda}_{xi}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i$ with an additional assumption that $E(\mathbf{x}_i) = \boldsymbol{\tau}_i + \mathbf{\Lambda}_{xi}E(\xi_i)$, where $E(\xi_i)$ is commonly denoted as $\boldsymbol{\kappa}_i$. The model with means requires zero means restrictions on the latent means parameters in the reference group (e.g., control treatment) in order to be identified and thus group or treatment means measure deviations from the reference group means. Sörbom's means-structure model assumes factorial invariance (i.e., invariance of the measurement model) across different groups. In the typical experimental design applications this assumptions precludes the effect of particular treatments on the inter-relationships among variables allowing merely different effects on means. This type of *ceteris paribus* assumption is thus problematic in a relatively large class of experimental treatment applications.

The multigroup estimation allows for more detailed analysis than generally required in experimental forestry research. Namely, it allows analysing in more detail covariance structures within each treatment group as well as testing for structural differences across groups. A relevant use of multigroup estimation for our purpose is to check whether the model from which we computed the latent scores holds approximately in each treatment group, but it additionally allows us to investigate treatment effects on the covariance structure of the measurements and their inter-relationships which might differ both across treatments and across time. An additional advantage of the multigroup approach is that this approach can

also solve the sample dependence and standard error problem of latent variable scores pointed out by Kano (2001).

4 Results

The covariance matrix for the entire sample is given in Table 3 with means in the bottom row. An informal look at the (panel) covariance matrix in Table 3 leads to an immediate conclusion that individual variances among the tree indicators differ to large degree despite the identical measurement scales (in centimetres). Covariances are also different, confirming the visual inspection of plots in Figure 1. Again, height in both years has far larger variance then the other two variables and it also correlates with them more. Height increase and diameter, on the other hand, correlate between themselves than with height individually. These findings indicate the greatest overall potential of height as a treatment discriminator.

Table 3: Panel data covariance matrix, $N = 1931$.

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1321.97					
x_2	311.36	136.66				
x_3	22.90	6.53	0.81			
x_4	1397.52	385.29	27.61	3030.98		
x_5	227.31	61.67	4.57	454.91	181.81	
x_6	27.37	8.11	0.66	50.23	9.79	1.44
$\overline{x}$	102.23	16.03	2.86	126.73	13.34	3.54

We first estimate the model specified in Eq. (2). Initially we estimate the basic panel model with uncorrelated errors. The $\mathbf{\Theta}_\delta$ matrix for this model is specified, writing only diagonal and lower triangular elements due to symmetry, as

$$\mathbf{\Theta}_\delta^{\mathrm{A}} = \begin{pmatrix} \theta_{11} & & & & & \\ 0 & \theta_{22} & & & & \\ 0 & 0 & \theta_{33} & & & \\ 0 & 0 & 0 & \theta_{44} & & \\ 0 & 0 & 0 & 0 & \theta_{55} & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} \end{pmatrix} \tag{7}$$

Estimation of this model gives an overall-fit χ^2 statistic of 158.72 (d.f. = 8). The *goodness-of-fit index* (GFI) is 0.97; the *standardised root-mean-square residual* (RMR) is 0.026; the *non-normed fit index* (NNFI) is 0.97; and the *root-*

mean-square error of approximation (RMSEA) is 0.098. Relaxing the cross-time error covariance parameters (θ_{41}, θ_{52}, θ_{63}) and estimating a modified model with $\boldsymbol{\Theta}_\delta$ specified as

$$\boldsymbol{\Theta}_\delta^B = \begin{pmatrix} \theta_{11} & & & & & \\ 0 & \theta_{22} & & & & \\ 0 & 0 & \theta_{33} & & & \\ \theta_{41} & 0 & 0 & \theta_{44} & & \\ 0 & \theta_{52} & 0 & 0 & \theta_{55} & \\ 0 & 0 & \theta_{63} & 0 & 0 & \theta_{66} \end{pmatrix} \qquad (8)$$

gave a χ^2 of 27.90 (d.f. = 5), GFI = 1.00; RMR = 0.016; NNFI = 0.99; CFI = 1.00; and RMSEA = 0.048, which jointly indicate approximately well fitting model.

Table 4: Treatment effects (N = 1931).

	1999 data							
	x_1	(S.E.)	x_2	(S.E.)	x_3	(S.E.)	ξ_1	(S.E.)
T_1	12.64	(2.85)	4.69	(0.91)	0.39	(0.07)	13.59	(2.65)
T_2	9.96	(2.88)	2.67	(0.92)	0.47	(0.07)	11.43	(2.67)
T_3	18.95	(2.77)	3.94	(0.89)	0.50	(0.07)	18.42	(2.58)
T_4	12.66	(2.67)	1.91	(0.86)	0.37	(0.06)	11.93	(2.48)
T_5	1.05	(3.01)	0.85	(0.97)	0.24	(0.07)	3.28	(2.80)
$F_{(5,1925)}$	13.04	–	7.47	–	15.27	–	13.29	–
	2001 data							
	x_4	(S.E.)	x_5	(S.E.)	x_6	(S.E.)	ξ_2	(S.E.)
T_1	22.15	(4.13)	3.48	(1.03)	0.59	(0.09)	22.61	(4.34)
T_2	14.91	(4.17)	2.58	(1.04)	0.45	(0.10)	16.15	(4.38)
T_3	26.74	(4.02)	1.98	(1.00)	0.56	(0.09)	27.65	(4.23)
T_4	18.17	(3.87)	0.74	(0.96)	0.47	(0.09)	18.51	(4.06)
T_5	–0.45	(4.37)	–2.63	(1.09)	0.20	(0.10)	–1.84	(4.59)
$F_{(5,1925)}$	14.51	–	7.51	–	11.98	–	14.62	–

T_1 = AV-SPS1; T_2 = AV-SPS2; T_3 = Biosol; T_4 = AV+Biosol; T_5 = Biomag

Table 4 shows dummy variable regression results for each of the tree variables in both years, together with the results for the latent variables, $\boldsymbol{\xi}_i$, calculated using estimated parameters from the model with $\boldsymbol{\Theta}_\delta$ specified as in Eq. (8). Note that the coefficient estimates express treatment effects as increase over the control (the omitted variable). Testing for treatment effects on the individual tree development indicators suggests that treatment T_4 (Biosol) has largest effect on tree height both in 1999 (x_1) and in 2001 (x_4), while T_2 (AV-SPS1) has relatively largest effect on

growth increase in both years (x_2 and x_5). Regarding the effects on trunk diameter (x_3 and x_6) there is a slight difference between the two years (though not statistically significant) were relatively best performance show T_4 (Biosol) in 1999 and T_2 (AV-SPS1) in 2001. The last column of Table 4 shows the results for the latent variable scores in 1999 (ξ_1) and in 2001 (ξ_2). There is no longer any ambiguity regarding the effects across years.

Table 5: Covariance matrices for the treatment sub-samples.

	Control (N = 362)						AV-SPS1 (N = 302)					
	x_1	x_2	x_3	x_4	X_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1467.62						1496.50					
x_2	323.46	148.98					326.45	117.30				
x_3	28.95	8.28	0.80				25.01	5.95	0.72			
x_4	1652.20	430.71	35.83	3188.79			1565.67	371.91	23.77	2849.15		
x_5	276.34	82.49	7.07	573.55	181.2		242.32	67.28	4.06	401.73	165.22	
x_6	31.74	9.08	0.90	65.88	13.01	1.73	31.48	8.25	0.53	52.19	8.92	1.39
$\bar{x}$	92.77	13.71	2.53	112.76	12.25	3.17	105.40	18.40	2.92	134.90	15.73	3.75

	AV-SPS2 (N = 292)						Biosol (N = 334)					
	x_1	x_2	x_3	x_4	X_5	x_6	x_1	x_2	X_3	x_4	x_5	x_6
x_1	1172.67						1159.77					
x_2	265.83	147.70					315.10	146.08				
x_3	12.60	4.54	1.20				20.90	5.70	0.64			
x_4	1407.82	457.91	25.46	2606.65			1202.21	356.35	23.35	2617.24		
x_5	246.24	93.50	4.16	505.10	187.13		203.70	82.76	3.54	384.95	174.63	
x_6	26.83	9.66	0.62	55.04	10.66	1.35	21.84	6.19	0.50	45.36	8.11	1.21
$\bar{x}$	102.73	16.38	3.00	127.67	14.83	3.62	111.72	17.65	3.03	139.50	14.23	3.73

	AV+Biosol (N = 392)						Biomag (N = 249)					
	x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1203.53						1477.65					
x_2	298.02	112.25					344.83	155.04				
x_3	19.70	5.38	0.54				31.35	9.93	1.13			
x_4	1080.34	241.43	15.96	2811.15			1572.57	515.59	46.87	4420.74		
x_5	230.99	72.40	4.17	557.22	222.30		141.50	44.66	4.07	220.86	142.41	
x_6	17.95	4.90	0.38	49.80	13.01	1.27	38.95	12.33	1.19	26.66	2.32	1.76
$\bar{x}$	101.00	12.99	2.83	125.78	12.68	3.35	93.81	14.57	2.78	112.31	9.62	3.37

Since our full sample has 1931 cases, each subsample in multi-group estimation was sufficiently large to perform the estimation. The numbers of trees under various treatments differ due to unequal tree survival rate across treatments

while the initial numbers of planted trees were equal for all treatments (including control). The covariance matrices and means together with *N* (sample size) are shown in Table 5.

Table 6: Maximum likelihood estimates of the coefficients.

	Separate estimates												**Joint estimates**	
	Control		T_1		T_2		T_3		T_4		T_5			
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
λ_{11}	33.02	1.59	36.77	1.72	26.88	1.78	33.54	1.45	32.30	1.30	33.01	2.01	32.69	0.68
λ_{21}	9.78	0.54	8.89	0.52	9.59	0.63	9.41	0.57	8.94	0.44	10.50	0.65	9.48	0.23
λ_{31}	0.87	0.04	0.68	0.04	0.52	0.06	0.63	0.04	0.61	0.03	0.95	0.05	0.69	0.02
λ_{42}	53.87	2.18	48.36	2.48	51.05	2.16	48.78	2.30	46.23	2.11	35.46	4.07	48.85	1.04
λ_{52}	10.60	0.60	8.20	0.68	9.91	0.69	8.03	0.67	12.11	0.63	2.20	0.43	9.30	0.28
λ_{62}	1.23	0.05	1.08	0.05	1.08	0.05	0.93	0.05	1.07	0.04	0.85	0.08	1.04	0.02
ϕ_{21}	0.77	0.02	0.80	0.03	0.94	0.02	0.74	0.03	0.54	0.04	0.99	0.08	0.81	0.01
θ_{11}	350.07	35.24	145.56	38.53	449.59	50.71	394.40	36.47	124.39	28.17	387.42	49.07	249.26	17.47
θ_{22}	53.40	4.53	38.25	3.25	55.85	6.33	57.41	5.31	31.02	2.98	45.42	4.72	46.22	2.04
θ_{33}	0.06	0.02	0.26	0.03	0.93	0.08	0.25	0.02	0.18	0.02	0.23	0.04	0.33	0.01
θ_{44}	225.60	55.60	518.23	86.23	266.00	40.97	236.16	98.22	660.81	69.28	316.33	29.59	649.65	40.98
θ_{55}	69.13	5.62	96.99	8.44	89.07	7.49	109.08	8.94	75.76	6.52	137.54	12.05	95.54	3.48
θ_{66}	0.27	0.03	0.22	0.04	0.18	0.02	0.35	0.04	0.12	0.03	1.04	0.10	0.37	0.02
θ_{41}	228.27	34.38	155.33	40.92	128.78	32.48	−31.60	39.63	222.41	30.73	−59.99	70.56	101.60	18.60
θ_{52}	−1.17	3.50	4.93	3.92	5.88	4.73	24.97	4.94	8.95	3.00	−2.55	5.68	−6.54	1.83
θ_{63}	0.11	0.02	−0.01	0.02	0.07	0.02	0.09	0.02	0.05	0.01	0.07	0.04	0.08	0.01
χ^2	11.85	–	12.24	–	7.02	–	13.01	–	9.18	–	10.03	–	27.90	–
d.f.	5	–	5	–	5	–	5	–	5	–	5	–	5	–
N	362	–	302	–	292	–	334	–	392	–	249	–	1931	–

T_1 = AV-SPS1; T_2 = AV-SPS2; T_3 = Biosol; T_4 = AV+Biosol; T_5 = Biomag

Table 6 gives maximum likelihood estimates of the model parameters for treatment subsamples (columns 1-6) and the full estimates from the joint model (column 7). Inspection of the parameter estimates suggests that structural parameters (λ_{11}, λ_{21}, λ_{31}, λ_{42}, λ_{52}, λ_{62}, and ϕ_{21}) are similar across. The expectation that different treatments affect tree growth measures not only in respect to their mean and variance but also in respect to their covariances is tested in the most general form with the Box-M test. The Box-M χ^2 was 2677.24 with 105 degrees of freedom which rejects general equality of covariance matrices. Further testing within the specific model proceeds with separate estimation of all parameters for all groups. The results are shown in Table 6. This is completely unconstrained model and its overall χ^2 statistic was 63.32 with 30 degrees of freedom. The χ^2 values shown in the bottom raw of Table 6 are those that would be obtained when the model is separately estimated for each particular group separately (note that there the overall χ^2 statistic for the joint model cannot be obtained by summing up individual χ^2's). Constraining the lambda matrices (factor loadings) to be equal across groups produces a χ^2 of 593.15 with 60 d.f. (GFI = 0.78, RMR = 0.19). Overall, we conclude that treatments indeed have an effect on covariances among the tree measurement indicators, and not merely on their means.

Testing a SEM model with means structure (Eq. 6) due to inclusion of additional parameters (means of the latent variables) and data (means of the observed variables) produces somewhat different results from the multi-sample analysis without means reported above. Repeating the multi-group comparisons with means included and first testing equality of factor loadings, i.e., holding $\boldsymbol{\tau}_i$ and $\boldsymbol{\Lambda}_{xi}$ fixed across subsamples, produced a χ^2 fit statistic of 475.90 (d.f. = 70), GFI = 0.90, RMR = 0.17, NNFI = 0.95, RMSEA = 0.13, which is a significantly better fit from the above estimated model without means. This indicates that the addition of means itself improves the fit of the model. In addition, the means-structure model allows direct estimation of the latent means providing comparative, but not identical, results to those reported in Table 4. As such estimates are rather dependable on the assumption of parameter equality across sub-samples, interpretation of these results warrants caution. We additionally test the equality of latent variable means ($\boldsymbol{\kappa}_i$) by fixing them to be equal across subsamples, again holding $\boldsymbol{\tau}_i$ and $\boldsymbol{\Lambda}_{xi}$ fixed across subsamples. This produced a χ^2 fit statistic of 593.24 (d.f. = 80), GFI = 0.90, RMR = 0.16, NNFI = 0.95, RMSEA = 0.14, which is significantly worst fit from the previously estimated model. This indicates that latent means are significantly different across treatment subsamples. Note that testing for equality of $\boldsymbol{\kappa}_i$ across treatment subsamples in the multigroup context has the same purpose as performing ANOVA tests on latent variable scores (or equivalently Wald F test in dummy variable regression). Table 7 reports the estimated latent means using Sörbom's technique performed on the model with $\boldsymbol{\Theta}_\delta$ matrix specified by Eq. (8). It can be noticed that most of the estimates are close to those obtained on the basis of factor means (Table 4). Table 7 shows estimates from two models, one with only factor loadings constrained across sub-samples (C), and the other with both loadings and error covariances constrained to be equal (D). A noted difference relates to the latent mean estimates in the T_4 treatment (AV+Biosol), which does not agree with the dummy variable regression results (Table 4) on individual observed variables separately. This last finding indicates high degree of sensitivity of the multi-group means-structure modelling when experimental treatments affect not only means but also general covariance pattern of the observed variables.

Table 7: Estimates of latent means from means-structure model.

	$\xi_1^{(C)}$		$\xi_1^{(D)}$		$\xi_2^{(C)}$		$\xi_2^{(D)}$	
Treatments	Estimate	(S.E.)	Estimate	(S.E.)	Estimate	(S.E.)	Estimate	(S.E.)
T_1	16.08	(2.86)	16.50	(2.83)	24.46	(4.16)	24.57	(4.15)
T_2	13.42	(2.76)	14.02	(2.72)	15.93	(4.24)	16.28	(4.23)
T_3	20.25	(2.71)	19.88	(2.67)	25.36	(4.01)	24.83	(3.97)
T_4	7.39	(2.59)	8.13	(2.57)	7.97	(3.97)	8.19	(3.94)
T_5	5.25	(3.14)	6.78	(3.18)	1.50	(4.17)	3.31	(4.21)

T_1 = AV-SPS1; T_2 = AV-SPS2; T_3 = Biosol; T_4 = AV+Biosol; T_5 = Biomag

However, since the equality of $\mathbf{\Lambda}_i$ matrices across groups is rejected, the comparison of factor means cannot be formally done, because the latent variables are compounded in a different way in different groups. Therefore, the results from Table 7 are given mainly for illustrative purposes.

5 Discussion

In this paper we estimated a panel model with three tree-development indicators taken at two different points in time using a general structural equation model with latent variables. This approach enabled statistical evaluation of the model specification and provided empirical guidance to subsequent model modification. Two main approaches were applied: the latent scores technique in combination with the classical experimental methods (i.e. dummy variable regression), and the multigroup structural equation modelling. Scores for the latent tree development were computed and used in testing for treatment effects. The subsequent analysis included multigroup estimation that aimed at comparing model specification in individual treatment subsamples.

The main results from this paper show that multivariate methods can be successfully used in experimental design with good potential of providing less ambiguous and more substantively interpretable results. The main requirement for using methods described in this paper is multivariate normality of the tree development indicators and linearity in their bivariate relationships. Our approach to dealing with the empirically observed deviations from normality was to transform the data, for which purpose the normal scores technique was used. This technique proved satisfactory in dealing with both non-normality and non-linearity.

Advantages of the latent scores approach are in the ability to apply standard experimental techniques such as ANOVA and dummy variable regression that have clear interpretability and familiarity for the applied researchers. It is substantively appealing to have a measure of "underlying tree development" for each individual tree in the sample and to be able to graph or do other post-secondary analysis on these quantities, which is enabled by the latent scores approach. On the other hand, the main drawback of the latent scores approach is that replacement of unknown structural parameters by their estimates induces sample dependence in the estimated latent scores, which can affect the accuracy of standard errors and test statistics (Kano, 2001). This indicates an important direction and need for further research, especially in the direction of adjustment of standard errors and fit statistics. In addition to needed adjustment for sample dependence bias, further work in this direction might develop adjustments needed for multiple range comparison tests that are of high importance in experimental research. Kano (2001) proposed to use the MIMIC model instead of the latent

scores approach, as a possible solution to the sample dependence bias. The MIMIC approach indeed has several potential advantages such as being simpler to apply by practitioners; it does not have standard error or test statistics problems; and it leads to more powerful tests of treatment effects. However, the MIMIC approach induces a small factor loading bias and it does not allow application of standard experimental techniques. Subsequently, extensions to multiple comparison tests within the MIMIC framework are not straightforward and would be rather difficult. Thus, while the latent scores are relatively complex to compute they allow the use of standard experimental testing techniques that are more popular in the applied experimental research.

The multigroup structural equation modelling approach, on the other hand, enables computation of treatment effects as well as testing for similarity in covariance structures across treatment subsamples, but it has rather restrictive assumptions regarding invariance of the model parameters across the subsamples. This equal covariance structure requirement is problematic in a relatively large class of experimental treatment applications where the effect of treatment alone is expected to cause differences in e.g. factor loadings structures across treatment subsamples. A relevant use of multigroup estimation for our purpose is to check whether the model from which we computed the latent scores holds approximately in each treatment group. However, while a requirement for treatment-effect testing in the multigroup framework this is not required for computing latent variable scores. Nevertheless, the multigroup approach, similarly to the MIMIC approach, might be able to solve the sample dependence and standard error problem of latent variable scores pointed out by Kano (2001).

Further research in this direction could focus on application of multiple comparison procedures and on adjustment of the significance levels. Such adjustment might be needed because the latent-scores were obtained from a multivariate statistical model that has been build or modified already on inferential grounds. Additionally, alternative paths of model building and different estimation methods could be considered. Data with more pronounced time series dimension might be modelled by extensions of the presently widely used structural equation models such growth curve modelling that requires at least three time points (see Muthén *et al*., 2001). Most importantly, the data gathering process and decision of which tree development indicators need to be measured would be improved by knowing how each measure fits into the covariance structure giving raise to overall tree development.

Acknowledgements

We would like to thank Ettore Dal Farra, Branimir Prpić, Ivica Tikvić, and two anonymous referees for helpful comments and suggestions. This work was

supported by the Austrian Ministry of Forestry and the Croatian Ministry of Science and Technology.

References

[1] Adler, D. (1995): *Growth Modelling for Mixed Tropical Forests*. Oxford Forestry Institute-University of Oxford-Forestry Research Programme.

[2] Arney, J.D. (1985): A modeling strategy for the growth projection of managed stands. *Canadian Journal of Forestry Research*, **15**, 511-518.

[3] Bagozzi, R.P. (1977): Structural equation models in experimental research. *Journal of Marketing Research*, **14**, 271-284.

[4] Bagozzi, R.P. and Yi, Y. (1989): On the use of structural equation models in experimental design. *Journal of Marketing Research*, **26**, 271-284.

[5] Bollen, K.A. (1989): *Structural Equations with Latent Variables*. New York: Wiley.

[6] Bossel, H. and Krieger, H. (1994): Simulation of multi-species tropical forest dynamics using a vertically and horizontally structured model. *Forest Ecology and Management*, **69**, 123-144.

[7] Botkin, D.B. (1993): *Forest dynamics, an ecological model*. Oxford: Oxford University Press.

[8] Bowman, K.O. and Shenton, L.R. (1975): Omnibus Test Contours for Departures from Normality based on $\sqrt{b_1}$ and b_2. *Biometrika*, **62**, 243-50.

[9] Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S., and Polić, M. (2002a): Covariance Structure Analysis of Regional Development Data: An Application to Municipality Development Assessment. In V. Bade, H. Heinrich, and H.L. Brauser (Eds.): F*rom Industry to Advanced Services: Perspectives of European Metropolitan Region*. Dortmund: European Regional Science Association.

[10] Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S., and Polić, M. (2002b): A Multivariate Methodology for Modelling Regional Development in Croatia. *Croatian International Relations Review*, Vol. VIII, **25**, 35-52.

[11] Cziráky, D., Puljiz, J., Jurlin, K., Maleković S., and Polić M. (2003): An Econometric Model for Development Level Assessment with an Application to Municipality Development Classification. In P. Filipić, W. Barlett, and M. Bateman (Eds.): *Enterprise in Transition*, Proceedings from the Fifth International Conference "Enterprise in Transition". Split: University of Split.

[12] D'Agostino, R.B. (1986): Tests for Normal Distribution. In R.B. D'Agostino and M.A. Stephens (Eds.): *Goodness-of-fit Techniques*. New York: Marcel Dekker, 367-419.

[13] Dittmar, C., Zech, W., and Elling, W. (2002): Growth variations of Common beech (Fagus sylvatica L.) under different climatic and environmental conditions in Europe, a dendroecological study. *Forest Ecology and Management*, **5856** (forthcoming), 1-16.

[14] Doornik, J.A. and Hansen, H. (1994): *A Practical Test for Univariate and Multivariate Normality.* Discussion paper. Oxford: Nuffield College.

[15] Hochberg, Y. and Tamhane, A.C. (1987): *Multiple Comparison Procedures*. New York: John Wiley.

[16] Jöreskog, K.G. (1971): Simultaneous factor analysis in several populations. *Psychometrika*, **36**, 409-426.

[17] Jöreskog, K.G. (1973): A General Method for Estimating a Linear Structural Equation System. In A.S. Goldberger and O.D. Duncan (Eds.): *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

[18] Jöreskog, K.G. (1999): *Formulas for Skewness and Kurtosis*. Chicago: Scientific Software International (http://www.ssicentral.com/lisrel).

[19] Jöreskog, K.G. (2000): *Latent Variable Scores and Their Uses*. Chicago: Scientific Software International. (http://www.ssicentral.com/lisrel).

[20] Jöreskog, K.G. and Goldberger, A.S. (1975): Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, **70**, 631-639.

[21] Jöreskog, K.G., Sörbom, D. du Toit, S., and du Toit, M. (2000): *LISREL 8: New Statistical Features.* Chicago: Scientific Software International.

[22] Kano, Y. (2001): Structural equation modeling for experimental data. In R. Cudeck, R., S., du Toit, and Sörbom, D. (Eds.): *Structural Equation Modeling: Present and Future: A Festschrift in honor of Karl Jöreskog*. Chicago: Scientific Software International, 381-402.

[23] Kaplan, D. (2000): *Structural Equation Modelling: Foundations and Extensions*. Thousand Oaks: Sage Publications.

[24] Lawley, D.N. and Maxwell, A.E. (1971): Factor Analysis as a Statistical Method. 2nd ed. London, Butterworths.

[25] Mardia, K.V. (1980): Tests of Univariate and Multivariate Normality. In P.R. Krishnaiah (Ed.): *Handbook of Statistics*. Vol. **I**, Amsterdam: North Holland, 279-320.

[26] Miller, R.G. (1981): *Simultaneous Statistical Inference*. 2nd ed. New York: Springer-Verlag.

[27] Muthén, B.O., Brown, C.H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C., Wang, C.-P., Kellam, S.G., Carling, J.B., and Liao, J. (2001): General growth mixture modelling for randomised preventive interventions. *Biostatistics*, forthcoming.

[28] Shenton, L.R. and Bowman, K.O. (1977): A Bivariate Model for the Distribution of $\sqrt{b_1}$ and b_2. *Journal of the American Statistical Association*, **72**, 206-11.

[29] Šidák, Z. (1967): Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, **62**, 626-633.

[30] Sörbom, D. (1974): A General Method for Studying Difference in Factor Means and Factor Structures Between Groups. *British Journal of Mathematical and Statistical Psychology*, **27**, 229-239.

[31] Sörbom, D. (1981): Structural equation models with structured means. In K.G. Jöreskog, and H. Wold (Eds.): *Systems under Indirect Observation, Causality, Structure and Prediction*. Amsterdam: North-Holland.

[32] Vanclay, J.K. (1994): *Modelling Forest Growth and Yield. Application to Mixed Tropical Forests*. Wallingford: CAB International.

[33] Wensel, L.C., Meeschaert, W.J., and Biging, G.S. (1987): Tree height and diameter growth models for Northern California conifers. *Agric. Exp. Stn. Hilgardia*, **55**, 1-20.

[34] Winer, B.J., Brown, D.R., and Michels, K.M. (1991): *Statistical Procedures in Experimental Design*. 3rd ed. New York: McGraw-Hill.