Running Head: *Longitudinal data analysis with SEM*

Longitudinal data analysis with structural equations

Jesús Rosel[(1)] and Ian Plewis[(2)]

(1) Departamento de Psicología Evolutiva, E., S. y Metodología.
Universitat Jaume I, Castellón, Spain

(2) Centre for Longitudinal Studies, Institute of Education, University of London, United Kingdom

Dr. Jesús Rosel
Departamento de Psicología E, E, S y Metodología
Universitat Jaume I
Apdo. 224,
12080 Castellón,
SPAIN.

Tel: 964 72 93 30.
Fax: 964 72 93 49.
E-mail: rosel@psi.uji.es

Suprimit: ¶

Longitudinal data analysis with structural equations

Abstract

In this paper we review different structural equation models for the analysis of longitudinal data: (a) univariate models of observable variables, (b) multivariate models of observable variables, (c) models with latent variables, (d) models that are unconditioned or conditioned to other variables (depending on the variability of the independent variables: time-varying or time-invariant, and depending on the type of independent variables: of latent variables or of observable variables), (e) models with interaction of variables, (f) models with non-linear variables, (g) models with a constant, (h) with single level and multilevel measurement, and (i) other advances in SEM of longitudinal data (latent growth curve model, latent difference score, etc.).

We have paid more attention to the interaction of variables and to non-linear transformations of variables because they are not frequently used in empirical investigation. They do, however, offer interesting possibilities to researchers who wish to verify relations between the variables they obtain. Potential applications are described, with their advantages and disadvantages.

Suprimit: We

Suprimit: unconditioned

Suprimit: and

Suprimit: ¶

Longitudinal data analysis with structural equations

Since Jöreskog (1969), Keesling (1972) and Wiley (1973) first developed the statistical model of structural equations, it has become one of the most widely used techniques for analysing longitudinal data. So much is this the case that: (a) in most handbooks about longitudinal data analysis one or several chapters are devoted to exemplifying structural equation models (Amsel & Renninger, 1997; Bijleveld & van der Kamp, 1998; Collins & Horn, 1991; Collins & Sayer, 2001; Dwyer, Feinlieb & Hoffmeister, 1992; Fitzmaurice, Laird & Ware, 2004; Frees, 2004; Gottman, 1995; Little, Schnabel & Baumert, 2000; Plewis, 1985; Singer & Willet, 2003; von Eye & Clogg, 1994); (b) centres that provide training in longitudinal research offer courses in structural equation modelling (SEM); and (c) there is a steady increase in the number of journal articles in which this methodology is applied (Card & Little, 2007). An example of the importance of SEM applied to longitudinal studies is reflected by the two chapters that Jöreskog (1974, 1977) published on this topic 30 years ago.

In this article we will briefly review the different models applied to the analysis of longitudinal data by means of SEM. We keep the statistical theory down to a basic level, but this work is also intended to be a practical guide for researchers who have to analyse longitudinal data. For this reason, we largely avoid dwelling on the basics of SEM and it is therefore assumed that the reader has some prior knowledge about this subject matter.

One important point to be taken into account in any SEM model is that the researcher, on conducting his or her research, must ensure a proper integration of the following aspects: (a) a basic theory with the correctly formulated hypotheses so as to be able to check whether the data match the theory; (b) a correct research design, with a thorough study of the variables to be measured, the time between measurements, the number of measurements, the age or ages of the sample, the time the research lasted, and so forth; and (c) the statistical model of data analysis that is to be used as a method of confirming (or, should it be the case, rejecting) the hypotheses that have been posited (Collins, 2006; Embretson, 2007; Little, Bovaird & Slegers, 2006; Little, Preacher, Selig & Card, 2007; Ram & Grimm, 2007).

Before beginning any longitudinal study it is important to put forward hypotheses about the stability of the observable and latent variables, as well as the relations between them, in order to check whether: (a) the variances of the latent and observable variables

Suprimit: Un

Suprimit: importante aspecto a tener en cuenta en cualquier modelo SEM es que el investigador, cuando lleva a cabo su investigación, ha de tener bien articulados entre sí los siguientes aspectos

Suprimit: una teoría de base con las correspondientes hipótesis correctamente formuladas con el fin de comprobar si los datos se corresponden con la teoría,

Suprimit: un correcto diseño de investigación

Suprimit: con las variables a medir, el tiempo entre mediciones, el número de mediciones, la edad o edades de la muestra, la duración de la investigación, etc. bien estudiadas,

Suprimit: el modelo estadístico de análisis de datos que sirva de método de comprobación (o, en su caso, de refutación) de las hipótesis planteadas

Suprimit: Antes de comenzar cualquier estudio longitudinal es importante plantear hipótesis acerca de la estabilidad de las variables observables y latentes

Suprimit: así como de las relaciones entre las mismas comprobando si

Suprimit: ¶

are stable or differ over time; (b) the same is true of the measurements of both the observable and latent variables; (c) the loads, or coefficients, between the latent variables and their corresponding observable variables are equal at each time of measurement; (d) the internal correlations between latent and/or observable variables remain stable between different times of measurement; and (e) the effects between different times of measurement are equal or changing (Brown, 2006).

Suprimit: las varianzas de las variables latentes y observables son estables o difieren a lo largo del tiempo,

Suprimit: lo mismo respecto de las medias, tanto de las variables observables como de las latentes,

Suprimit: las cargas, o coeficientes, entre las variables latentes y sus correspondientes variables observables son iguales en cada momento de medición,

Suprimit: las correlaciones internas entre variables latentes y/o observables se mantienen estables entre diferentes momentos de medición,

Suprimit: los efectos entre diferentes momentos de medición son equivalentes o cambiantes

The different models we will be looking at are as follows: (a) univariate and multivariate models; (b) observable variable models and latent variable models; (c) unconditioned versus conditioned to other variables; (d) with and without interaction terms; (e) with linear and non-linear effects; (f) with and without constants; (g) with measures at one and more than one level; and (h) diverse advances in longitudinal SEM. In fact, there can be as many different models as there are combinations among the preceding conditions, which is what makes SEM a highly versatile research tool.

Univariate models of observable variables

The theoretical foundation of effect models with observable variables lies in simple regression (Pearson, 1896) and in path analysis (Wright, 1918, 1921). Path analyses are a further development of regression, but with the addition of intermediate variables to predict the response variable of interest. SEM encompasses both these observable variable models and those with latent variables (or factors).

Univariate repeated measures models are those in which the same persons are measured on a single variable on several occasions. Thus, if the variable V_1 is measured 4 times (e.g. let us suppose we are dealing with a group of children for whom the variable 'level of knowledge of mathematics' is measured at 6, 7, 8 and 9 years of age; in panel designs this is represented by *4W1V*, that is to say, 4 'waves' or times, and 1 variable), then the representation of the variable at the four times would be: $V_{1,1}$, $V_{1,2}$, $V_{1,3}$, and $V_{1,4}$, where the first subscript indicates the variable (in this example it is always the same: $V_{1,1}$) and the second subscript indicates the time of measurement (the first time, at the age of 6, is represented by $V_{1,1}$; the second time, at 7 years old, by $V_{1,2}$, and so forth). The simplest model for representing the nature of these data is shown in Figure 1.

Suprimit: ,

Insert Figure 1 about here

In the graphic representation of the SEM we have followed Bentler's (1995) system of notation, which uses rectangles to represent the observed variables ($V_{1,1}$, $V_{1,2}$, $V_{1,3}$ and $V_{1,4}$); the effects of some variables on others are represented by arrows (which show the

Suprimit: ¶

direction of each respective effect), with the corresponding size of the effect (b_1, b_2 and b_3); and the arrows that are labelled with the letter "*E*" ($E_{1,2}$, $E_{1,3}$ and $E_{1,4}$) are the residuals.

The model in Figure 1 is called a first-order autoregressive model (*AR(1)*), or 'simplex' or Markov model. The term 'autoregressive' refers to the fact that each value of the variable V_1 at time *t* depends only on the value of that same variable in the previous measurement *t-1* ($V_{1,t}=f(V_{1,(t-1)})$). This model was put forward by Guttman (1954), but has since been developed by others (Anderson, 1960; Heise, 1969; Humphreys, 1960). It should be noted that there is no $E_{1,1}$ in the model because one of the assumptions of the model is that the independent (or explanatory) variables have been measured without error. This model can be represented algebraically in a compact form:

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\varepsilon}, \qquad (1)$$

where $\mathbf{y}$ is the $p \times 1$-order vector of dependent (or response) variables: $\mathbf{y}' = [V_{1,2}, V_{1,3}, V_{1,4}]$, $\mathbf{x}$ is the $q \times 1$-order vector of independent variables: $\mathbf{x}=[V_{1,1}]$, $\boldsymbol{\varepsilon}$ is the vector of residuals: $\boldsymbol{\varepsilon}' =[E_{1,2}, E_{1,3}, E_{1,4}]$, while the relations between the dependent variables are represented by matrix $\mathbf{B}$, of order $p \times p$, and are as follows:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ b_2 & 0 & 0 \\ 0 & b_3 & 0 \end{bmatrix}, \qquad (2)$$

$\mathbf{\Gamma}$ is a coefficient matrix of order $p \times q$ that relates the independent variables with the dependent ones, which in our case will be: $\mathbf{\Gamma}' = [b_1, 0, 0]$.

In order to estimate the corresponding parameters, matrix $\mathbf{\Psi} = Cov(\boldsymbol{\varepsilon})$, where *Cov* is the covariance operator so that $Cov(Y,Y) = Var(Y)$, and $\mathbf{\Psi}$ is a matrix of order $p \times p$, must be added to the previous matrices. In our case, the residual variances ($Cov(E_{1,2}, E_{1,2})$, $Cov(E_{1,3}, E_{1,3})$, $Cov(E_{1,4}, E_{1,4})$) will be represented on the main diagonals of this matrix. Matrix $\mathbf{\Phi} = Cov(\mathbf{x})$, of order $q \times q$, which represents the matrix of the covariances of the independent variables, would also have to be added to the model.

The model in Figure 1 (which has 3 degrees of freedom) can be reformulated so that each measurement depends on the measurement immediately before it but also on the measurement performed at the last time but one. As a result, the model in Figure 2 could be proposed.

Suprimit: ¶

Insert Figure 2 about here

The model in Figure 2 (which would have 1 degree of freedom) is called second-order autoregressive (*AR(2)*) because each value is a function of the two measurements immediately preceding it ($V_{1,t}=f(V_{1,(t-1)},V_{1,(t-2)})$), and if there were evidence that the measurements of the different times were closely related, a third-order autoregressive model could be contemplated, where a new effect (b_6) would have to be added to the model in Figure 2. This effect would go from $V_{1,1}$ to $V_{1,4}$, and in general for any *AR(3)* model it would be: $V_{1,t}=f(V_{1,(t-1)},V_{1,(t-2)},V_{1,(t-3)})$. To be able to formulate a *p*-order autoregressive model, we need at least $p+1$ measurements.

The *AR(3)* model above would have *0* degrees of freedom, and therefore it would not be possible to estimate coefficients of overall fit for the model. Different hypotheses about the variables can be tested. The main purpose of these hypotheses (as they are shown in Figures 1 and 2) are the equality of the effect parameters, of the variances and between covariances of the residuals (Brown, 2006):

Suprimit: hypothesis

Suprimit: hypothesis

(a) If it is assumed that the effects among variables remain constant over time, they can be constrained so that coefficients continue to be equal; thus, in the *AR(1)* model in Figure 1, it could be hypothesised that $b_1 = b_2 = b_3$ and, following on with the same logic, in the *AR(2)* model in Figure 2, it could be considered that $b_1 = b_2 = b_3$, and also $b_4 = b_5$. These assumptions make more sense if the time span between variables is the same, and if it is assumed that the process does not change over time (which is more likely to occur with adults than with small children).

Suprimit:

Suprimit: :

Suprimit: :

Suprimit: :

(b) Since the same variable $V_{1,t}$ is measured on several occasions, the residual variances could be assumed to be equal; thus, in the models in Figures 1 and 2, $Cov(E_{1,2}, E_{1,2}) = Cov(E_{1,3}, E_{1,3}) = Cov(E_{1,4}, E_{1,4})$ can be included in the syntax of the constraints paragraph of the input for the statistical software that is being used (*Cov* is the covariance operator so that $Cov(Y,Y) = Var(Y)$). This assumption makes more sense when the variances of the variables have roughly the same value, which is more frequent in research conducted with adults. This is due to the fact that when the same variable is examined in small children, the variance is usually seen to increase with age and the previous constraint will not be fulfilled empirically.

(c) Different assumptions can be made about the residuals. It might be reasonable to suppose that, because the same variable is being measured several times, the residuals at time 2 ($E_{1,2}$) will covary with those from time 3 ($E_{1,3}$), that is to say, $Cov(E_{1,2},E_{1,3})$ will

Suprimit: ¶

be left free (i.e. ≠ 0), those from time 3 will covary with those from 4 ($Cov(E_{1,3},E_{1,4})$), which could also be left free, and so forth, if there were more measurement times ($Cov(E_{1,4}, E_{1,5})$, …). These parameters will therefore be included in the corresponding covariance section as being free, since by default they are assumed to be equal to *0*. If this assumption were added to Figure 1, the resulting model would be like the one in Figure 3. From Figure 3 onwards, we will follow the convention of representing covariances by means of curved lines with arrowheads at both ends.

Suprimit: ,

Suprimit: ,

Insert Figure 3 about here

(d) Bearing in mind that it is the same variable, and that the same covariance processes occurring between errors can be repeated between consecutive measurements, the following constraint can be added: $Cov(E_{1,2}, E_{1,3}) = Cov(E_{1,3}, E_{1,4}) = etc$.

(e) The residuals at each occasion can be interpreted as 'innovations' of the variable at the time of measurement (Box, Jenkins and Reinsel, 1994), as they are a part of the same variable that is not explained by the previous variable or variables. Thus, a first-order moving average model ($MA(1)$) can be put forward, in which each value of the variable at any given time ($V_{1,t}$) is a function of the previous error for the same person ($E_{1,t-1}$), that is $V_{1,t} = f(E_{1,t-1})$, which results in the model in Figure 4.

Suprimit: ing

Insert Figure 4 about here

In fact, the model in Figure 4 is a first-order autoregressive and first-order moving average model ($ARMA(1,1)$). It is possible to add the constraint whereby the effects of the moving averages are invariant over time (as is assumed in Box-Jenkins time series models), by leaving $b_4 = b_5$. The constraints mentioned above for the effects, variances and covariances of the residuals can also be applied to the model in Figure 4.

The foregoing statistical hypothesis must be based on a theoretical justification for the model, but if used wisely they help to identify the underlying model that generates the data. The models that have been outlined here constitute the basis of SEM, and understanding and being able to apply them is an almost indispensable condition before moving on to more complex models.

Suprimit: hypotesis

Multivariate models of observable variables

Multivariate models of observable variables, as we have already stated, are an extension of univariate models. For example, let us suppose that (as in Figure 1) the same children are measured at 6, 7, 8 and 9 years of age on the variable 'level of knowledge of

Suprimit: ¶

language' in addition to their 'level of knowledge of mathematics': the design would now be represented by *4W2V*. As we have seen, the variable 'level of knowledge of mathematics' was represented at the four times by $V_{1,1}$, $V_{1,2}$, $V_{1,3}$ and $V_{1,4}$ and now the 'level of knowledge of language' is represented as $V_{2,1}$, $V_{2,2}$, $V_{2,3}$ and $V_{2,4}$. One advantage of measuring two variables is that it becomes possible to study the directional effect of one variable on the other. That is to say, the researcher can formulate two research questions, namely, do the two variables develop independently of each other, or does the variable 'level of knowledge of mathematics' exert a greater influence over that concerning 'level of knowledge of language' or vice versa? This gives us the model in Figure 5.

Suprimit: ;

Insert Figure 5 about here

In Figure 5 the covariance between the variables $V_{1,1}$, $V_{2,1}$ is represented by lines with arrowheads at both ends that join the two variables; if the variable 'maths ability' ($V_{1,1}$, $V_{1,2}$, $V_{1,3}$ and $V_{1,4}$) exerted an influence on 'language ability' ($V_{2,1}$, $V_{2,2}$, $V_{2,3}$ and $V_{2,4}$), then the effects b_8, b_{10} and b_{12} could be expected to be statistically significant; if the opposite were true, then the significant effects would be b_9, b_{11} and b_{13}. In each case the researcher must have some substantive hypotheses that justify the specification; if only a few of the previous effects were significant, the researcher would have to establish an explanation for these findings and their meaning.

The mathematical representation of the model in Figure 5 would have the same form as the one in Equation 1, but the **y** vector of the dependent variables is now of order *6×1*: **y'** = ($V_{1,2}$, $V_{1,3}$, $V_{1,4}$, $V_{2,2}$, $V_{2,3}$, $V_{2,4}$), the **x** vector is now of order *2×1*: **x'** = ($V_{1,1}$, $V_{2,1}$), while **ε** is the vector of residuals, which in our case is of order *6×1*: **ε'** = ($E_{1,2}$, $E_{1,3}$, $E_{1,4}$, $E_{2,2}$, $E_{2,3}$, $E_{2,4}$), **B** being of order *6×6*:

Suprimit: ($V_{1,2}$ $V_{1,3}$ $V_{1,4}$ $V_{2,2}$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ b_2 & b_{11} & 0 & 0 & 0 & 0 \\ b_{10} & b_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_3 & b_{13} & 0 & 0 \\ 0 & 0 & b_{12} & b_6 & 0 & 0 \end{pmatrix},$$

and **Γ** is a coefficient matrix of order *6×2* that relates the independent variables with the dependent ones:

Suprimit: ¶

$$\mathbf{\Gamma'} = \begin{pmatrix} b_1 & b_8 & 0 & 0 & 0 & 0 \\ b_9 & b_4 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Matrix **Ψ** is of order *6×6*, and in our case the variances of the residuals (*Cov*($E_{1,2}$, $E_{1,2}$), *Cov*($E_{1,3}$, $E_{1,3}$), …, *Cov*($E_{2,4}$, $E_{2,4}$)) will be represented on its main diagonal; the main diagonal of matrix **Φ** = *Cov*(**x**), of order *2×2,* will contain the variances of $V_{1,1}$ and of $V_{2,1}$, while the other two elements of the matrix contain the covariance between $V_{1,1}$ and $V_{2,1}$.

Sometimes the crossed effects between variables $V_{1,t}$ and $V_{2,t}$ are contemporaneous (either reciprocal or unidirectional) rather than lagged; thus, it can be supposed that in the model in Figure 5 the variables 'level of knowledge of mathematics' and 'level of knowledge of language' influence each other at the same time. The model would therefore be non-recursive (two or more variables exert an influence on each other); a feasible representation of such a model is shown in Figure 6.

Insert Figure 6 about here

If the variable 'level of knowledge of mathematics' ($V_{1,1}$, $V_{1,2}$, $V_{1,3}$ and $V_{1,4}$) influenced 'level of knowledge of language' ($V_{2,1}$, $V_{2,2}$, $V_{2,3}$ and $V_{2,4}$), then only effects b_9, b_{11} and b_{13} in Figure 6 would be significant. If, in contrast, the variable 'level of knowledge of language' were the one that exerted an influence on 'level of knowledge of mathematics', then parameters b_8, b_{10} and b_{12} would be significant. If *no* crossed effects between the variables (b_8, b_9, b_{10}, … b_{13}) were significant, then this would indicate that the variables $V_{1,t}$ and $V_{2,t}$ develop independently, with just the initial covariance (*Cov*($V_{1,1}$, $V_{2,1}$)) representing the preliminary relation between the two.

Suprimit: by

Suprimit:

The model in Figure 6 is non-recursive because there are one or several interdependence 'loops' between variables; note that there is a reciprocal influence between $V_{1,2}$ and $V_{2,2}$, between $V_{1,3}$ and $V_{2,3}$, as well as between $V_{1,4}$ and $V_{2,4}$. These models are also called interdependence and simultaneous equation models and can lead to identification difficulties (Bentler & Raykov, 2000; Bollen, 1989; Hayduk, 1996).

The constraints outlined in the section about univariate models of observable variables for the effects, variances and covariances can be added to the models in Figures 5 and 6, but in Figures 5 and 6 the covariances of contemporaneous measurements can be taken as being free: *Cov*($E_{1,2}$, $E_{2,2}$), *Cov*($E_{1,3}$, $E_{2,3}$), *Cov*($E_{1,4}$, $E_{2,4}$). In Figures 5 and 6, this could be represented by lines with arrowheads at both ends that join the respective measurement errors. The constraint *Cov*($E_{1,2}$, $E_{2,2}$) = *Cov*($E_{1,3}$, $E_{2,3}$) = *Cov*($E_{1,4}$, $E_{2,4}$)

Suprimit: ¶

could still be used in Figures 5 and 6 to add the assumption that the three contemporaneous covariances have the same value.

Models with latent variables

Models with latent variables (also called factors) suppose the existence of a variable that is not measured directly, but which determines the values of the observable variables that are manifestations of that same latent variable. Thus, for a particular person, the empirical value of an observable variable is a function of the score for the latent variable obtained by that individual and the measurement error. In this way, if we have several measurements carried out in different mathematics tests and a latent variable is established for those same scores, the grade obtained by a child in a test is a function of the one attained in the factor, plus their corresponding error. Spearman (1904) was the author who developed the factorial analysis model but, as has been stated above, SEM embrace regression, path analysis and factorial analysis. It can be hypothesised that the variables measured in Figure 1 are determined by one factor, and the model would now therefore be the one shown in Figure 7. Note that, following the convention usually employed in SEM, the latent variables are represented inside ovals or circles.

Insert Figure 7 about here

This model would correspond to the equation:

$$\mathbf{y} = \mathbf{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad (4)$$

where $\mathbf{y'} = [V_{1,1}, V_{1,2}, V_{1,3}, V_{1,4}]$, but now there are no longer values for $\mathbf{x}$, but instead scores for the latent variable F_1; $\boldsymbol{\varepsilon}$ is the measurement error vector: $\boldsymbol{\varepsilon'} = [E_{1,1}, E_{1,2}, E_{1,3}, E_{1,4}]$, while the effects of the latent variable F_1 on the dependent variables will be: $\mathbf{\Lambda'}_y = [b_1, b_2, b_3, b_4]$, and matrix $\boldsymbol{\eta} = [F_1]$.

Notice how this model indicates that the values obtained for each person on the measurement variable $V_{1,t}$ are a function of the value obtained for each child on the latent variable F_1. Note too that, bearing in mind that we are dealing with time data, each value of $V_{1,t}$ can be expected to be a function of F_1, but also of $V_{1,(t-1)}$ (that is to say, $V_{1,t} = f(F_1, V_{1,(t-1)})$). If this hypothesis were formulated, the resulting model would be as shown in Figure 8.

Insert Figure 8 about here

Because the model in Figure 8 is saturated, for it to be estimated (if it is assumed that the autoregressive effect is the same over time) the parameters can be set by means of

Suprimit: ¶

the following constraint: $b_5 = b_6 = b_7$. Alternatively, it can be assumed that the effect of the latent variable on the observed variables is the same over time: $b_1 = b_2 = b_3 = b_4$. The same constraints concerning the variances and the covariances of the errors can be applied to the models in Figures 7 and 8 as those applied in the section 'Univariate models of observable variables'.

Note that the estimation of the values of F_1 and of the parameters b_1, b_2, b_3 and b_4 are related, so that if the variance of F_1 is multiplied by 2, the values of the parameters b_1, b_2, b_3 and b_4 would be divided by 2 to give the same solution (and the same would occur with any other value, since on multiplying the variance of the factor, the value of the corresponding coefficients would automatically be divided), and the system would therefore admit an infinite number of solutions. This problem in factorial analysis is solved by standardising the values of F_1, that is to say, by setting the mean of F_1 to *0* and the variance to *1*. In SEM the most frequent solutions consist in either setting the variance of F_1 to *1* or making one of the coefficients of F_1 on its observed variables equal to unity (in this case, b_1=*1* is usually set); another solution is to make equal to one the mean of the effects of the latent variable on their own observed variables (Little, Slegers & Card, 2006). Note that SEM, by default, differentiate variables with regard to their respective mean; they do not therefore estimate constants and only model the relation between variances and covariances of the variables.

Suprimit: doing equal to one

Suprimit: equal to one

It may be thought that the observed variable, $V_{1,1}$, $V_{1,2}$, $V_{1,3}$, $V_{1,4}$, is the measure with error at each time of measurement implicit in its corresponding latent variable; the model would thus now be that shown in Figure 9 (Jöreskog, 1974, 1977). The measurement errors of the factors (or disturbances) are represented by means of the letter *D* with the subindex that corresponds to the factor.

Suprimit: ment

Suprimit: measure

Insert Figure 9 about here

Remember that the metrics have to be set for each factor; it is therefore advisable to set the values of each coefficient b_4, b_5, b_6 and b_7 as being equal to unity, the metrics for each factor being equivalent to those of their corresponding observable variable.

This system of estimation has several advantages, above all when the data obtained do not have the same time interval. But let us suppose (in the opposite case) that data about the variable $V_{1,t}$ had been measured at *6*, *7*, *9* and *10* years of age for each child (only the measurement at the age of 8 is missing). One way of representing and estimating the variable development model would be that shown in Figure 10, in which factor F_3

Suprimit: ¶

would represent the 'missing' measurement at the age of 8 (Dunn, Everitt & Pickless, 1993). Note that the effect of F_2 on F_4 is multiplicative (in standardised scores: $b_2 \times b_3$) and not additive, but the values of b_2 and b_3 are usually smaller than one in standardized values.

Suprimit: normally
Suprimit: that
Suprimit: 1

Insert Figure 10 about here

In Figure 10 each value of the effects b_5, b_6, b_7 and b_8 must also be set to equal unity, while the effects among latent variables can be left as being equal ($b_1 = b_2 = b_3 = b_4$), if it is assumed that the effects among factorial scores are the same over time.

SEM mainly model the covariances and variances between the different variables and factors (especially if no constants are added to the model), but there are cases in which the variance of the variable measured over time increases or diminishes systematically. A good system, therefore, for modelling that gradual change consists in using latent variables of a temporal nature that take into account the variability in the variance. Figure 11 shows a 'random walk' (or Wiener) model in the time factors.

Insert Figure 11 about here

As has been pointed out earlier, in measuring cognitive variables or those concerning the performance of small children, the dependent variable usually increases with age and so the variability at measurement time 1 is determined by F_1, and the increase in variability at each age is given by the influence of the successive factors F_2, F_3 and F_4. It is wise to set all the effects of the factors on the variables (from b_1 to b_{10}) as having a value of one; by so doing it is assumed that the effect of the variability of each time factor is kept constant across the different variables over time and that each successive factor increases the variability of the measurements in a constant manner. This model can also be reformed with additional hypotheses about autoregressive effects between factors, or between observable variables, and constraints on the values of the variances or covariances of the errors can be added.

From what has been outlined so far about models that have latent variables with a single (univariate) variable measured at different times, it is easy to generalise the procedure for multivariate models (in which several variables are measured at each time). If two variables were measured at each time, it would be possible to put forward the hypothesis that the measurements at each time are a function of one factor (i.e. if two performance variables were measured, the factorial capacity of the child at each time

Suprimit: ¶

determines the values obtained on the observable variables). Moreover, the value of each factor is influenced by that of the factor immediately before it and would result in the model shown in Figure 12.

Insert Figure 12 about here

It could be hypothesised that the autoregressive coefficients between the factors are equal ($b_1 = b_2 = b_3$) and even that the effects of the factors on the respective observed variables remain constant over time (of the factors over $V_{1,t}$: $b_4 = b_6 = b_8 = b_{10}$, and those of the factors over $V_{2,t}$: $b_5 = b_7 = b_9 = b_{11}$). As far as the variance of the errors is concerned, another hypothesis could be formulated by which the variances of each respective measurement error belonging to the same variable are equal, that is to say, with respect to the variable $V_{1,t}$: $Cov(E_{1,1}, E_{1,1}) = Cov(E_{1,2}, E_{1,2}) = Cov(E_{1,3}, E_{1,3}) = Cov(E_{1,4}, E_{1,4})$, and/or alternatively, with respect to $V_{2,t}$: $Cov(E_{2,1}, E_{2,1}) = Cov(E_{2,2}, E_{2,2}) = \ldots = Cov(E_{2,4}, E_{2,4})$.

With regard to the covariance among these errors, the covariance between observable variables for the same measurement time can be left free, since it is to be expected that an individual who had a 'bad' day at one time of measurement would have lower scores for that time, and that if he or she had a 'good' day then the scores for both variables would be higher. Thus, $Cov(E_{1,1}, E_{2,1})$, $Cov(E_{1,2}, E_{2,2})$, … $Cov(E_{1,4}, E_{2,4})$ would be left free. It is even logical to suppose that the previous covariances would be equal. Likewise, the covariances between consecutive times can be left free, but within the same variable; that is to say, for the variable $V_{1,t}$: $Cov(E_{1,1}, E_{1,2})$, $Cov(E_{1,2}, E_{1,3})$ and $Cov(E_{1,3}, E_{1,4})$, and for $V_{2,t}$: $Cov(E_{2,1}, E_{2,2})$, $Cov(E_{2,2}, E_{2,3})$, and $Cov(E_{2,3}, E_{2,4})$. Because we are dealing with a repetition of the same variable, the equality of the covariance values can be included in the constraints paragraph. Thus, for $V_{1,t}$: $Cov(E_{1,1}, E_{1,2}) = Cov(E_{1,2}, E_{1,3}) = Cov(E_{1,3}, E_{1,4})$, the same procedure being followed for the covariances between the errors for $V_{2,t}$.

Suprimit:

The previous model can be expanded by developing a system in which there is an autoregressive model in the time measurement factors, and with factors indicating the variables. In Figure 13 the time measurement factors would be F_1, F_2, F_3 and F_4, and the factors indicating the variables $V_{1,t}$ and $V_{2,t}$ are F_5 and F_6, respectively.

Insert Figure 13 about here

Suprimit: ¶

Thus, if any observable variable is taken, its measurement depends on the time factor and on the factor indicating the variable, which for the values of $V_{1,t}$ will be factor F_5, whereas for those of $V_{2,t}$ it will be F_6. The covariance between the factors indicating the variables ($Cov(F_5,F_6)$) is taken because it is assumed that both variables measured are of the same type; if we were to hypothesise that the factors indicating the variables $V_{1,t}$ and $V_{2,t}$ (F_5 and F_6) would covary with the first time of measurement (F_1), then $Cov(F_1,F_5)$ and $Cov(F_1,F_6)$ would have to be left free. It will also be necessary for one of the coefficients of each factor indicating the variables to be given a value of unity (for example: b_{12} and b_{16}); in this case it can be assumed that the effect of each indicating factor on its corresponding variables is the same. If this were true, then we could confirm the hypothesis about the effect between F_5 and its corresponding variables: $b_{13} = b_{14} = b_{15}$, and between F_6 and its variables: $b_{17} = b_{18} = b_{19}$. The same constraints concerning the variances and covariances as those stated earlier can be applied to this model.

If we formulate the hypothesis that there is a second-order factor that influences the time measurements and the factors indicating the respective variables, then it becomes possible to create a new latent variable F_7, which will exert an influence on: F_1, F_2, …, F_6 (in this case we would also have to add the measurement errors that were previously independent variables, D_1, D_5 and D_6, and establish a value of unity for any effect that went from F_7 to one of the other factors). This second-order factor (F_7) would indicate that the variables measured are very stable and that people change while maintaining the same relative position over time in each variable, so that those who maintain a high score in one of the variables at the first measurement of the first variable tend to have a high score at the first measurement of the second variable – a pattern that is repeated over time (Rosel & Elósegui, 1991).

As was pointed out earlier, univariate models of observable variables are the basis for understanding SEM applied to longitudinal designs, and multivariate models are an extension of univariate designs. Latent variable models, however, represent an important leap forward in terms of quality with respect to the previous section and are an extremely versatile and useful tool with which to study change in processes that take place in the human, biological and social sciences.

Suprimit: :

Suprimit: and biological

Unconditioned models versus those conditioned to other variables

Suprimit: ¶

The SEM we have discussed so far are unconditioned because we only measure the same dependent variable or the same dependent variables several times, the independent variable being the first measurement of the variable. Thus, the same variable was independent or dependent according to the time of measurement.

It is possible to formulate models that are conditioned to other variables by making these the independent (or exogenous) variables inherent to the system itself; these independent variables can admit two variations, which are: (i) depending on the variability of the independent variables: time-varying or time-invariant (i.e. that do or do not change over time, respectively), and (ii) depending on the type of independent variables: latent variables or observable variables.

So, if in Figure 5 it is assumed that there were a latent variable 'general cognitive level' (F_1) that was measured at the first time and that this were determined by two variables, that is, age (X_1) and the number of correct answers on Raven's progressive matrix test (X_2), F_1 in this case being the independent variable of the system, the model would now be the one shown in Figure 14, which would have a latent variable (F_1) that is invariant over time.

Insert Figure 14 about here

The model in Figure 14 cannot be estimated directly with the LISREL program because no relation between the observable dependent variables is allowed by this statistical formulation (when they derive from an observable variable that indicates a latent variable; that is to say, the LISREL model does not admit relationships between $V_{1,1}$ and $V_{2,1}$ or towards the remaining variables); in contrast, it can be estimated with other programs (such as EQS, for example). One way of estimating the model in Figure 14 with the LISREL program is by using 'phantom variables' (Rindskopf, 1984) (i.e., each dependent observable variable, $V_{1,1}$, $V_{1,2}$, $V_{1,3}$, $V_{1,4}$, $V_{2,1}$, $V_{2,2}$, $V_{2,3}$, $V_{2,4}$, is identified by means of a latent variable without error and with an effect on the corresponding observable variable equal to one.

Instead of using a latent variable as the 'motor' of the system, observable variables can also be used. Thus, if the independent variables X_1 and X_2 do not form a factor, but they are hypothesised as being the exogenous variables of $V_{1,t}$ and $V_{2,t}$, then the model in Figure 15 can be proposed.

Insert Figure 15 about here

Suprimit: that

Suprimit: El

Suprimit: modelo de la Figura 14 no puede ser estimado directamente con el programa LISREL, puesto que las variables dependientes observables (cuando parten de una variable observable que es indicadora de una variable latente; es decir, el modelo LISREL no admite relaciones desde $V_{1,1}$ o $V_{2,1}$ hacia el resto de variables: bien entre ellas mismas o hacia el resto de variables) no pueden tener relaciones entre sí; en cambio, sí puede ser estimado con otros programas (como p.ej.: el EQS)

Suprimit: Una

Suprimit: to make sure that

Suprimit: Thus, the manera de estimar el modelo de la Figura 14 con el programa LISREL es hacer que cada variable observable dependiente: $V_{1,1}$, $V_{1,2}$, $V_{1,3}$, $V_{1,4}$, $V_{2,1}$, $V_{2,2}$, $V_{2,3}$, $V_{2,4}$ esté identificada mediante una variable latente sin error, y con efecto sobre la correspondiente observable latente igual a la unidad, así, las variables $V_{1,1}$, $V_{1,2}$, $V_{1,3}$ y and $V_{1,4}$, por on the one hand, and the un lado, y las variables $V_{2,1}$, $V_{2,2}$, $V_{2,3}$ y and $V_{2,4}$, on the otherpor el otro lado, would be left with a configuration similar to the one in Figure 9quedarían con una configuración semejante a la de la Figura 9, although now there would be no aunque ahora no habría disturbances (without D's). By so doing, the metric of each observable variable and its corresponding latent variable would be identical De este modo, la métrica de cada variable observable y su correspondiente latente sería idéntica,. As a result, the effects among latent variables would be the same as those obtained using another program (for example, EQS) and, moreover, the level of fit of the model and the degrees of freedom coincide con lo cual, los efectos entre las variables latentes serían los mismos que los obtenidos mediante otro programa (p.ej.: el EQS), además, el nivel de ajuste del modelo y los grados de libertad coinciden.

Suprimit: ¶

As in the model in Figure 14, if we propose that X_1 and X_2 present a permanent effect over all the measurement times, the corresponding effects of X_1 and X_2 on $V_{1,2}$, $V_{1,3}$, $V_{1,4}$, $V_{2,2}$, $V_{2,3}$ and $V_{2,4}$ can be added.

The two time-invariant conditioned models described here, that is, latent variables (see Figure 14) and observable variables (see Figure 15), can be generalised and applied to the models of latent variables mentioned above, for example, from Figure 8 to Figure 13.

With respect to the time-varying conditional models, let us suppose the model in Figure 12, in which the cognitive capacity of each child has been measured at each time as the number of correct answers in an intelligence test ($X_{1,t}$). This exerts an influence at each respective measurement time and, through the respective factors, would affect their capacity for language and mathematics; the result would therefore be the model in Figure 16.

Insert Figure 16 about here

Note that in Figure 16 we have included all the possible covariances between the independent variables $X_{1,1}$, $X_{1,2}$, $X_{1,3}$ and $X_{1,4}$: $Cov(X_{1,1}, X_{1,2})$, $Cov(X_{1,1}, X_{1,3})$, ... $Cov(X_{1,3}, X_{1,4})$, assuming that there is a correlation between the children's scores at each time of measurement. If we assumed that the independent variable $X_{1,t}$ was temporally dependent on itself (instead of covariation), then we would have to remove the covariances between $X_{1,1}$, $X_{1,2}$, $X_{1,3}$ and $X_{1,4}$ and put in the effects that correspond from $X_{1,1}$ to $X_{1,2}$, from $X_{1,2}$ to $X_{1,3}$ and from $X_{1,3}$ to $X_{1,4}$.

Suprimit:

Figure 17 shows a development model conditioned to time-varying factors. If we assume that the following are measured in the children: the variable 'level of knowledge of mathematics', represented at the four times by $V_{1,1}$, $V_{1,2}$, $V_{1,3}$, and $V_{1,4}$; the variable 'level of knowledge of language', represented as $V_{2,1}$, $V_{2,2}$, $V_{2,3}$, and $V_{2,4}$; and a general intelligence test ($X_{1,t}$) and Wechsler's block design test ($X_{2,t}$) are used as independent variables (or covariables), and that both intelligence tests form a temporal intelligence factor (F_5, F_6, F_7 and F_8) and, furthermore, that there is a time dependence between the factors, then the model would be the one in Figure 17.

Insert Figure 17 about here

Note that, statistically, the independent variable of the system is factor F_5, which means it has been measured with no error; if we were to assume that the temporal intelligence factor (F_5, F_6, F_7 and F_8) is covarying over time (instead of having a temporal

Suprimit: ¶

dependence), then we would have to remove effects b_{16}, b_{17} and b_{18}, the errors of measurement of the factors D_6, D_7 and D_8, and add the covariances among factors F_5, F_6, F_7 and F_8: $Cov(F_5, F_6)$, $Cov(F_5, F_7)$, ..., $Cov(F_7, F_8)$.

The same considerations as those made in the section about 'Univariate models of observable variables' and 'Multivariate models of observable variables' can be repeated here with respect to the measurement errors of the variables $V_{1,t}$ and $V_{2,t}$ in order to gain degrees of freedom and facilitate the convergence of the model estimators.

As can be seen, including substantive ('time-invariant' or 'time-varying') independent variables in SEM adds new possibilities to the estimation of effects between variables, and to the search for dependence among them.

Models with interaction of variables

When two or more variables are being measured, one of them, or a new dependent variable, can be hypothesised as being a function of the interaction of two (or more) independent variables (Aiken & West, 1991; Aiken, West & Pitts, 2003; Jaccard, Turrisi & Wan, 2003). Interaction of variables is understood to mean the product of them. Thus, if we propose that the interaction between consuming alcohol and tobacco favours the appearance of cancer, then we have to measure the rate of alcohol and tobacco consumption for each person, and we must create a third variable that will be the product of both for each person, while at the same time, as a dependent variable, it would serve as an indicator of cancer. In interaction of variables, an overriding principle is that of 'nesting': when an interaction of variables is used, we must utilise the variables and corresponding interactions that lie on a lower level. Thus, if the interaction of the variables *X, Y* and *Z* were used in a model, the variables and interactions *XYZ, XY, XZ, YZ, X, Y, and Z* would have to be employed so that the model makes substantive sense and the forecasts established by the model can be performed correctly.

There are three interaction procedures in SEM: interaction between independent observable variables, interaction when the observed variables are part of the same factor and interaction between factors, with their corresponding observable variables (Jaccard, Turrisi & Wan, 1999; Kenny & Judd, 1984; Ping, 1995; Schumacker, 2002; Schumacker & Marcoulides, 1998; Schumacker & Lomax, 2004).

Suprimit:

Suprimit: ¶

Suprimit: rate

Suprimit: ¶

Let us suppose that in Figure 5 the interaction of $V_{1,t}$ and $V_{2,t}$ exerts an influence on the development of the variables $V_{1,t}$ and $V_{2,t}$. Thus, if a new variable is generated, it should be the interaction of $V_{1,t}$ and $V_{2,t}$ for each time:

$$V_{1*2,t} = V_{1,t} * V_{2,t}, \quad (1)$$

In this way we would have a new variable for each measurement time and, bearing in mind that it has an influence on $V_{1,t}$ and $V_{2,t}$, we would propose the model in Figure 18.

Suprimit: in

Insert Figure 18 about here

Note that in Figure 18: (a) the interaction between variables $V_{1*2,1}$, $V_{1*2,2}$ and $V_{1*2,3}$ have been included in the model; (b) the covariances between all the independent variables of the model have also been included; (c) the variable $V_{1,2}$ is a function of $V_{1,1}$, of $V_{2,1}$ and of the interaction between both of them ($V_{1*2,1}$). The same thing happens with the variables $V_{1,3}$ [$V_{1,3} = f(V_{1,2}, V_{2,2}, V_{1*2,2})$] and $V_{1,4}$ [$V_{1,4} = f(V_{1,3}, V_{2,3}, V_{1*2,3})$], and the variable $V_{2,t}$ depends on the interaction of $V_{1*2,t-1}$, that is to say: $V_{2,2} = f(V_{1,1}, V_{2,1}, V_{1*2,1})$, $V_{2,3} = f(V_{1,2}, V_{2,2}, V_{1*2,2})$] and $V_{2,4} = f(V_{1,3}, V_{2,3}, V_{1*2,3})$.

It can also be supposed that the interaction between variables is time dependent and, in that case, we would need to remove $Cov(V_{1*2,1}, V_{1*2,2}, V_{1*2,3})$ and add both $V_{1*2,2} = f(V_{1*2,1})$ and $V_{1*2,3} = f(V_{1*2,2})$.

Suprimit:

Should there be an interaction between variables within the same factor, so that they affect another factor, we would have to create a new factor resulting from the interaction between the two of them. Thus, if in Figure 12 it is assumed that the interaction of the variables $V_{1,t}$ and $V_{2,t}$ within each factor F_t affects the factor F_{t+1}, we would have to obtain the product of the variables from equation (6), including the product within a new factor, as shown in Figure 19.

Insert Figure 19 about here

We can hypothesise that F_5, F_6 and F_7 are time dependent, and therefore we would need to remove $Cov(F_5, F_6, F_7)$ and add $F_6 = f(F_5)$ and $F_7 = f(F_6)$.

Suprimit: ,

It is easy to apply the interaction of variables between factors and the references cited above can be consulted for further information on this matter.

In models of interaction of variables it is also possible to use the same procedure as that proposed in relation to the errors in measuring the variables $V_{1,t}$ and $V_{2,t}$ in the previous sections.

Suprimit: ¶

Models with non-linear variables

Strictly speaking, these ought to be called models with transformation to non-linear variables; thus, it becomes possible to propose a model in which a variable (Y_t) is a function of the value of the same variable at a previous time (Y_{t-1}) and of the square of the same variable at the previous time (Y^2_{t-1}). In other words, $Y_t=f(Y_{t-1},Y^2_{t-1})$, which indicates that the relation between Y_t and Y_{t-1} is quadratic. This non-linear relation can be of any other type (and not only quadratic): cubic, quadratic, logarithmic, square rooted, and so forth, or other more complex, exponential-type functions: logistic, Gomperz, and so forth. Even interaction models are a special case of non-linear transformation (Etezadi-Amoli & McDonald, 1983; Jaccard, Turrisi & Wan, 1999; Jöreskog & Yang, 1996; Kenny & Judd, 1984; McArdle & Nesselroade, 2003; Schumacker, 2002; Schumacker & Marcoulides, 1998; Seber & Wild, 1989).

If, in the model in Figure 1, it were assumed that $Y_t=f(Y_{t-1},Y^2_{t-1})$, then the graphic representation would be that shown in Figure 20.

Insert Figure 20 about here

In Figure 20: (a) the squares of the respective independent variables have been included in the model; (b) the covariances between all the independent variables in the model are also included, $Cov(V_{1,1}, V^2_{1,1}, V^2_{1,2}, V^2_{1,3})$ meaning all the pairwise covariances of the independent variables, that is to say, $Cov(V_{1,1}, V^2_{1,1}, V^2_{1,2}, V^2_{1,3})$: $Cov(V_{1,1},V^2_{1,1})$, $Cov(V_{1,1},V^2_{1,2})$, ..., $Cov(V^2_{1,2},V^2_{1,3})$; and (c) note that the variable $V_{1,2}$ is a function of $V_{1,1}$, and of the square of the same variable [$V_{1,2}=f(V_{1,1},V^2_{1,1})$]. The same thing happens with the variable $V_{1,3}$, which is a function of $V_{1,2}$ and of the square of this same variable ($V^2_{1,2}$) (that is: $V_{1,3}=f(V_{1,2},V^2_{1,2})$), and lastly: $V_{1,4}=f(V_{1,3},V^2_{1,3})$.

If the quadratic effects were not significant, it would be necessary to withdraw effects b_4, b_5 and b_6, together with their corresponding independent variables ($V^2_{1,1}$, $V^2_{1,2}$ and $V^2_{1,3}$). By so doing, the model would become the one shown in Figure 1 again.

Instead of the quadratic model between observable variables suggested here, any other non-linear transformation of the variables could also be hypothesised.

The non-linear model of observed variables can be generalised to latent variables and, thus, a quadratic model of the effect of one factor (F^2_t) on another factor (F_{t+1}) could also be proposed and, hence, F_{t+1}=f(F_t, F^2_t). The model in Figure 9 would then become the one in Figure 21.

Suprimit: :

Suprimit: ¶

Insert Figure 21 about here

Each quadratic factor must contain its respective squared variables and the corresponding interaction of the variables. It is strongly advisable to include orthogonal coefficients in the product and the power of the variables in order to avoid colinearity (Little, Bovaird & Widaman, 2006); this is also valid for interaction between variables, as we have explained in the previous section.

The covariances among the pairwise independent variables, $Cov(F_1, F_5, F_6, F_7) = Cov(F_1, F_5), Cov(F_1, F_6), ..., Cov(F_6, F_7)$, must be added to Figure 21.

Supposing the quadratic factors were not significant, then the corresponding effects b_{12}, b_{13} and b_{14} would not be significant either. If the quadratic effect b_{13} were significant (that is to say, from F_6 to F_3), but one of the effects of F_6 on its indicator variables were not significant, then all the effects and their corresponding indicator variables would have to be left (due to the principle of hierarchy, for the appropriate forecast of F_3).

We can also assume that F_5, F_6 and F_7 are time dependent and, therefore, we would have to remove $Cov(F_5, F_6, F_7)$ and add $F_6 = f(F_5)$ and $F_7 = f(F_6)$.

When the time intervals between successive measurements are different and it is also assumed that the passage of time exerts a constant effect on the different measures, the following strategy can be used to include non-linear effects among the variables: (i) each measuring unit is identified by a constant (e.g. a); (ii) each autoregressive coefficient is made equal to the constant a raised to the power of the time interval; additionally, (iii) the value of a must be less than unity, so that the more time has passed, the lower the effect of the passage of time will be. Thus, let us suppose in Figure 21 that one year passes between the first and the second measurement, a year and a half between the second and the third, and 10 months between the third and the fourth measurement. In the constraints paragraph of the syntax of an SEM statistical program it could be indicated (by reducing the time unit to months) that $b_1 = a^{12}$, $b_2 = a^{18}$, $b_3 = a^{10}$, and $a<1$.

The criteria of the variances and covariances of the errors that were proposed earlier can be applied.

Both the procedures of interaction between variables and those involving their non-linear transformation enable the researcher to optimise the information that is available. This is because, although generally speaking data collection tends to be the most

Suprimit: In factor F_5 (factor of F_1 squared), the parameters have to be set in accordance with the following criterion: $b_{15}=(b_4)^2$; $b_{16}=b_4*b_5$; $b_{17}=(b_5)^2$; $b_{18}=(b_6)^2$; $b_{19}=b_6*b_7$; $b_{20}=(b_7)^2$; $b_{21}=(b_8)^2$; $b_{22}=b_8*b_9$; $b_{23}=(b_9)^2$; that is to say, the variable $V_{1.1}$ squared ($V^2_{1.1}$) that receives the effect of F_5 must have the square of the coefficient of F_1 on $V_{1.1}$, the same being true for the factors F_6 and F_7 with respect to their variables.

Suprimit: El

Suprimit: The system for estimating parameters for models that include interactions of variables or non-linear effects can be implemented in two phasessistema de estimación de parámetros para modelos que incluyan interacción de variables o efectos no-lineales puede ser implementado en dos fases (Ping, 1995), and

Suprimit: even possible

Suprimit: e incluso pueden incluirse coeficientes ortogonales en el producto de parámetros para evitar la colinealidad

Suprimit: over

Suprimit: para

Suprimit: incluir efectos no-lineales entre las

Suprimit: ¶

demanding part, with the data already available to the researcher it is possible to test more hypotheses concerning the same variables with their corresponding transformation.

Models with constant

SEM initially relate covariances and variances between the variables. The model differentiates (or centres) the observable and latent variables with respect to the mean, with the resulting loss of information about the constants. Nevertheless, it is sometimes very important to obtain the value of the constants of the model. In longitudinal designs, this is the case when comparing the results in terms of an intervention in different groups or to determine the change of level in a variable.

As an example, let us suppose that in the case of Figure 1 we were interested in verifying the constant at each measurement time; this would therefore give us the model in Figure 22.

Insert Figure 22 about here

In the representation of the constants we followed the system put forward by McArdle and Epstein (1987), in which the constants are distinguished by an arrow with a triangle placed at the origin with a number '1' inside it. It must be borne in mind that the variable $V_{1,1}$ now admits forecast error ($E_{1,1}$). The substantive interpretation of the other elements and symbols in Figure 22 is the same as that for Figure 1 (except in the constants that have been added).

The model in Figure 22 can easily be extended by including the vector of constants in equation (1):

$$\mathbf{y} = \boldsymbol{\alpha}_y + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\varepsilon}, \qquad (6)$$

where additionally: $\mathbf{x} = \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$.

Interpretation of the symbols is the same as that explained in Equation 1. In our case, $\boldsymbol{\alpha}_y' = [k_1\ k_2\ k_3\ k_4]$, $\boldsymbol{\alpha}_x$, $\boldsymbol{\Lambda}_x$, $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are null matrices, and $\boldsymbol{\Psi}' = [(Cov(E_{1,1}, E_{1,1}), (Cov(E_{1,2}, E_{1,2}), Cov(E_{1,3}, E_{1,3}), Cov(E_{1,4}, E_{1,4})]$. Similarly, the substantive interpretation of the other elements and symbols in Figure 22 is the same as that for Figure 1 (except in the constants that have been added).

One important aspect to be taken into account in Figure 22 is that, as in the other regression models, the values of the constants do not necessarily have to be those of the means (Rosel, Arnau & Jara, 1998; Rosel, Jara & Arnau, 2002). Instead, they are

Suprimit:)];
Suprimit: similarly
Suprimit: ¶

conditioned to the value of the mean of the variable that they have as a dependent variable, and also that of the effects they receive through the independent variables. For example, in Figure 1, let us suppose that the means of each variable were $\overline{V}_{1,1} = 50$, $\overline{V}_{1,2} = 60$, $\overline{V}_{1,3} = 70$ and $\overline{V}_{1,4} = 80$ and that the results of the coefficients were $b_1 = 0.60$, $b_2 = 0.70$ and $b_3 = 0.85$; the equation of the variable $V_{1,1}$ would be $V_{1,1} = k_1 + E_{1,1}$, and its expected values $E(V_{1,1}) = E(k_1 + E_{1,1})$. As a result, $\overline{V}_{1,1} = k_1$, and the value of k_1 would therefore be the mean of $V_{1,1}$, that is, $k_1 = 50$; for $V_{1,2}$ it would be $V_{1,2} = k_2 + b_1 \cdot V_{1,1} + E_{1,2}$, and its expected values would be $E(V_{1,2}) = E(k_2 + b_1 \cdot V_{1,1} + E_{1,2})$, which results in $\overline{V}_{1,2} = k_2 + b_1 \cdot \overline{V}_{1,1}$, $k_2 = \overline{V}_{1,2} - b_1 \cdot \overline{V}_{1,1} = 60 - 0.60 \cdot 50 = 30$. Hence, the value of k_2 no longer depends only on the mean of $V_{1,2}$, but also on the coefficient b_1 and on the mean of $V_{1,1}$. Likewise, we would have $k_3 = 28$, and $k_4 = 20.50$, which proves that the value of the constants diminishes over time, despite the fact that the means of the variables increase. As can be seen with this simple example, there does not necessarily have to be a relation between the means and the constants, and therefore the magnitudes of the constants cannot be taken directly to compare the means.

In conclusion, the values of what most authors call 'structural means' are not these means at all but instead constants from each respective forecast equation. We recommend comparing the means of the observable variables taking the covariances among variables (with no effects among them) as the reference model, which thus allows the respective means to be compared directly. On the other hand, for the latent variables it is advisable to use the procedure proposed by Little, Slegers and Card (2006), which constrains the mean of the effects of the latent variable on its corresponding observable variables to a value of one.

With single level versus multilevel measurement

Up to this point the models have been developed under the assumption that the parameters are fixed. In other words, we have not considered the possibility that these parameters (for example: b_1, in Figure 1) might vary across units, perhaps from one higher level unit such as a geographical area to another. Suppose we have a response or dependent variable, V, measured on four possibly unequally spaced occasions with any pattern of missing data after the first occasion. Let us also suppose that our sample is structured in such a way that each case is located in one and only one cluster (pupils (i) within schools (j), for example). Our observations are V_{tij}, t (measurement occasion) =

Suprimit: :
Suprimit: :
Suprimit: :
Suprimit: :
Suprimit: Como
Suprimit: conclusión
Suprimit: los valores de lo que la mayoría de autores denomina 'medias estructurales', no son tales medias, sino que son constantes de cada respectiva ecuación de pronóstico
Suprimit: Nuestra recomendación es comparar las medias de las variables observables estableciendo como modelo de referencia las covarianzas entre las variables (sin ningún efecto entre las mismas),
Suprimit: así se comparan directamente las respectivas medias;
Suprimit: mientras para las medias de las variables latentes es conveniente usar el procedimiento de
Suprimit: que constriñe igual a uno la media de los efectos de la variable latente sobre sus correspondientes variables observables
Suprimit: ,
Suprimit: ¶

1, ..., T_{ij} (max. T_{ij} =*4*); *i* (case) = *1*, ..., n_j; *j* (cluster) = *1*, ..., *J* and we will assume that $\mathbf{V} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

We can approach the analysis of these data in a number of ways, depending on our underlying research question. The first method is to model V_t as a function of one or more earlier or lagged measures of *V*, that is V_{t-k}, $k > 0$, along with other explanatory variables. This approach is particularly useful if we seek causal conclusions about, for example, the effect of an intervention in the absence of randomisation (see, for example, Plewis, 1985). We can regard each of the three responses V_t, $t > 1$, as a multivariate set (V_1 is assumed to be exogenous here) and so our model is:

$$V_{2ij} = b_{20j} + b_{21}V_{1ij} + e_{2ij} \quad (2.a)$$

$$V_{3ij} = b_{30j} + b_{31}V_{1ij} + b_{32}V_{2ij} + e_{3ij} \quad (2.b)$$

$$V_{4ij} = b_{40j} + b_{41}V_{1ij} + b_{42}V_{2ij} + b_{43}V_{3ij} + e_{4ij} \quad (2.c)$$

$$b_{k0j} = b_{k00} + u_{kj}, k = 1,2,3 \quad (2.d)$$

The intercept terms (b_{k0j} , $k = 2, 3, 4$) vary randomly from cluster to cluster around an overall mean b_{k00} as shown in equation (2.d). The regression coefficients b_{k1} represent the relations between each response and the exogenous V_1; b_{k2}, $k = 3, 4$ link the last two responses to their value on the second occasion and b_{43} links the final response to its previous or lagged value.

Equations (2.a) to (2.c) define a recursive model for the cases, equation (2.d) defines the model at the cluster level and we assume that the case residuals are uncorrelated with the cluster residuals. Maximum likelihood estimation is relatively straightforward, albeit iterative, and can be carried out in most statistical packages as well as with the more specialised multilevel packages such as *MLwiN* and *HLM*. Providing any missing data are 'missing at random', in other words, any missing V_t ($t > 1$) depend only on V_{t-k}, $k > 0$, then the algorithms are fully efficient.

The model generates two covariance matrices, one for the cases and the other for the clusters. The multivariate model allows all the residuals at the case level to be correlated as might happen if there is an omitted explanatory variable that is related to V_t ($t > 1$) after controlling for V_{t-k}, $k > 0$, in other words, to the change in *V* from occasion to occasion. This part of the model is essentially the same as the model for 'seemingly unrelated regressions' as introduced into the econometrics literature by Zellner (1962).

The model can be simplified. For example, we might constrain the parameters b_{31}, b_{41} and b_{42} to be zero so that any response V_t depends only on its immediately preceding

Suprimit: ...,
Suprimit: ...,
Suprimit: ...,
Suprimit: i.e.
Suprimit: ,
Suprimit: exogeneous
Suprimit: exogeneous
Suprimit: at
Suprimit: specialist
Suprimit: ;
Suprimit: for
Suprimit: ¶

measure as in Figure 1 for a single level model. In the same vein, we might also constrain some of the elements of the covariance matrix at both levels to be zero so that only the first-order autoregressive elements are non-zero. On the other hand, it will sometimes be appropriate to make the model more complicated. In particular, we might want to allow some or all of the regression coefficients b to vary across clusters so that we replace b_{pq} by b_{pqj} and the covariance matrix at the cluster level is extended to include both the extra variances but also the covariances between the intercepts (b_{k0j}) and the slopes (b_{pqj}). We might also want to introduce explanatory variables measured at the cluster level and these might include cluster variables created from the explanatory variables at the case level, such as the cluster mean.

Rather than relating the response at one occasion to lagged measures of the response, a second approach to analysing repeated measures data comes from modelling the evolution of the response over time (or age). In other words, we model V_t as a function of time within a growth curve or unconditional framework. This can be particularly useful in a developmental or ageing context. Our model is now:

$$V_{tij} = \sum_{q=0}^{Q} \beta_{qij} a_{tij}^{q} + e_{tij} \qquad (3.a)$$

$$\beta_{qij} = \beta_{q0j} + u_{qj} \qquad (3.b)$$

$$\beta_{q0j} = \beta_{q00} + v_{qj} \qquad (3.c)$$

Here Q is the order of the polynomial at age (or time) a, determined by the data but with the restriction that $Q << max\ T_j$, and β_q are the model coefficients that show how the response varies with age. These (random) coefficients can vary from case to case and from cluster to cluster.

The equations (3.a) to (3.c) are a three-level model with variation between occasions within cases (level one), represented by σ_e^2, variation between cases within clusters (level two) represented by the random effects ***u*** for the intercept, slope etc., and variation between clusters (level three) represented by a second set of random effects ***v***. It can be estimated with the same algorithm used for model 3, with equations (3.a) to (3.c). In the structural equations literature, the random effects ***u*** and ***v*** are sometimes treated as latent variables that are determined by the observed responses over time (Meredith & Tisak, 1984, 1990).

The multilevel growth curve model is a flexible tool that can be adapted and extended. Thus, it can accommodate the fact that there is often variation in the age or time of

Suprimit: in

Suprimit: *a*

Suprimit: three

Suprimit: and

Suprimit: ¶

measurement within any one measurement occasion. It is possible to allow the level-one residuals to be autocorrelated (Goldstein et al., 1994) and to allow the variance at level one to change with age or time. Explanatory variables can be introduced at each level (time-varying variables at level one, case-level variables at level two and cluster variables at level three) and these variables can be used to explain variation in any of the growth parameters β_q.

The conditional (or regression) and the unconditional (or growth curve) approaches can, with sufficient occasions of measurement, be combined, as shown by Plewis (1996). Then it is possible to condition on the measure at the first occasion (the baseline measure), thereby making it possible to compare growth parameters across groups for cases starting off at the same point. Other extensions are also possible – for example, to binary, unordered and ordered categorical responses (Plewis et al., 2006), to multivariate responses (Plewis, 2005), and to combinations of these.

Suprimit: –

Suprimit: –

Other advances in SEM of longitudinal data

Although we do not intend to go into great depth on the matter, other advances in SEM of longitudinal data have been made in recent years, some of the most important being the following:

(a) Latent growth curve models (Bollen & Curran, 2006; Duncan, Duncan & Strycker, 2006; Meredith & Tisak, 1984, 1990), which use latent variables and constants. One of the latent variables would thus represent the starting level and the other latent variable would represent the rate of growth of the group, but now adding indicator constants to each latent variable. New latent variables that represent the quadratic, cubic or some other tendency of the development curve can also be added. In this model the respective variances of the latent variables indicate the random variability of the individuals that make up the sample, and therefore each latent variable becomes a variable with a random (multilevel) coefficient. This latent curve model is thus directly related to the multilevel model. Figure 23 shows an example of a linear latent growth curve model with a random constant and a linear effect. Note that all the coefficients from the latent variables towards the observable variables are fixed effects and that F_1 represents the level of the constant (fixing the values: $b_1 = b_2 = b_3 = b_4 = 1$), while F_2 is the linear slope of the model (fixing the values: $b_5 = 1$, $b_6 = 2$, $b_7 = 3$).

Insert Figure 23 about here

Suprimit: En

Suprimit: la Figura

Suprimit: se muestra un ejemplo de modelo de

Suprimit: con una constante aleatoria y un efecto lineal

Suprimit: Obsérvese

Suprimit: que todos los coeficientes desde las variables latentes hacia las variables observables son efectos fijos (fixed effects) y que

Suprimit: representa

Suprimit: el nivel de la constante

Suprimit: mientras

Suprimit: es

Suprimit: la pendiente lineal del modelo

Suprimit: ¶

(b) Another model that is useful for explaining development is that known as the latent difference score model (McArdle, 2001), shown in Figure 23, which has the following components: (i) some first-order latent variables, which represent the true measure of each observable variable, (ii) a second-order latent variable at each time of measurement (ΔF_t, or latent difference score), which would represent the gain in latent scores ($\Delta F_t = F_t - F_{t-1}$), (iii) a third-order latent variable, which would exert an influence over each second-order latent variable, and would denote an overall gain factor for each person (or slope), and (iv) a constant that affects the initial second-order and third-order factor. Thus, it is found that the change in the true scores at each time depends on: (1) the initial level of each person, (2) the overall gain (slope), (3) the scores on the same variable at the previous time, and (4) if we are dealing with a multivariate model, the true score of the other variables at the previous time. If the variance of the initial level and the overall gain are seen to be significant, then we would have two random values for each person, and this model is therefore also related with the random coefficients model (the article by Ferrer, Balluerka and Widaman (2008) in this journal can be consulted for more complex second-order latent growth curve and latent difference score models). Figure 24 shows a univariate model of the latent difference score model, and should include some constraints on the parameters: $b_1 = b_2 = \ldots = b_9 = b_{10} = b_{17} = 1$, and $b_{11} = b_{12} = b_{13}$ (proportional change) and $b_{14} = b_{15} = b_{16}$ (additive change). It is advisable to perform more than two measurements (variables) each time. The advantage of the difference score model is that it can also be applied to cross-sectional data.l artículo de Ferrer, Balluerka8

Insert Figure 24 about here

(c) One system of analysis also developed within the context of longitudinal designs in SEM is the one described by Nagin (1999, 2005) and by Muthén and Shedden (1999; Muthén, 2001), which classifies individuals in clusters, according to a probability of assignment to the group. This procedure is called the 'growth mixture model', and sorts individuals into classes according to patterns of development profile. The drawback with this procedure is that it is an exploratory system, and must therefore be used with caution.

(d) Since the mediational model of effects among variables in cross-sectional designs was developed (Judd & Kenny, 1981; Baron & Kenny, 1986), several attempts have been made to adapt the procedure to longitudinal designs. One of the most interesting

Suprimit: que tiene los siguientes componentes

Suprimit: unas variables latentes de primer orden, que representan la medida verdadera de cada variable observable

Suprimit: En la Figura

Suprimit: se muestra un modelo univariado de

Suprimit: ;

Suprimit: en la Figura 24

Suprimit: habría de ponerse algunas constricciones a los parámetros

Suprimit: ...

Suprimit: La ventaja del modelo de

Suprimit: es que también se puede aplicar a datos transversales

Suprimit: (The article by Puede verse en esta misma revista e

Suprimit: and Widaman (200

Suprimit:) in this journal can be consulted for more complex second-order latent growth curve and latent difference score modelspara modelos más complejos de segundo orden de latent growth curve y de latent difference score).

Suprimit: Desde

Suprimit: que desarrollaron el modelo de mediación de efectos entre variables en diseños transversales (

Suprimit:)

Suprimit: cross-sectional

Suprimit: varios intentos se han hecho para adaptar el procedimiento a diseños longitudinales

Suprimit: ¶

approaches is the one put forward by Cole & Maxwell (2003), who make a distinction between mediation and direct effects in panel designs, and outline a series of steps to be used to check whether real mediation exists among the variables under study.

(e) Steyer (2005; Steyer, Eid & Schwenkmezger, 1997; Steyer, Krambeer & Hannöver, 2004) proposed a latent variable model to determine the values of both individual and overall change; this model is similar to the latent difference score model, but it is easier to parametrise. It is wise to collect at least two observed variables at each measurement time.

Suprimit: Una de las aproximaciones más interesantes es la de

Suprimit: &

Suprimit: and

Suprimit: que distingue entre mediación y efectos directos o indirectos en diseños en panel

Suprimit: proporcionando una serie de pasos para comprobar si existe mediación real entre las variables de estudio

Suprimit: partly inspired

Suprimit: in

Suprimit: requires

Discussion

Following this quick overview of SEM applied to longitudinal designs, we might be lulled into thinking that it is easy to fit a model to the data obtained, but nothing could be further from the truth. It is difficult to make a model fit the data properly but when we do manage to do so, explaining the data becomes a simple matter.

In this review, we have paid more attention to interaction between variables and to non-linear transformation of variables because they are not frequently used in empirical research. They do, however, offer interesting possibilities to researchers who wish to verify relations between the variables they obtain.

Suprimit: lineal

Suprimit: investigation

The researcher has to put forward several foregoing hypotheses about his or her data and then look for the model that best fits them rather than trying out an array of models (mining the data) with nothing to offer guidance about their nature. Some authors, however, following on from Tukey (1977), argue that statistical theory must be adapted to the needs of scientific method and that the search for a new statistical model capable of generating real data is an unavoidable stage in the processes of constructing scientific theories and models that explain the data (Marcoulides & Drezner, 2001).

To statistically verify which model is the best among nested models, we can use the χ^2 comparison procedure proposed by Jöreskog (1974, 1979) or, if the models are not nested, the values from the Bayesian information criterion (or BIC) can be used (Raftery, 1995; Schwartz, 1978).

Suprimit: For

Suprimit: ing

In recent years the number of publications that utilise SEM in data analysis has risen sharply and this percentage is very likely to remain steady and even increase, due to the enormous possibilities offered by this method for analysis and the ease with which it can be adapted to different data collection circumstances.

Suprimit: ¶

References

Aiken, L.S. & West, S.G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA.: Sage.

Aiken, L.S., West, S.G. & Pitts, S.C. (2003). Multiple linear regression. In: J.A. Schinka & N.F. Velicer (Eds.), *Research Methods in Psychology*. New Jersey: John Wiley & Sons.

Amsel, E. & Renninger, K.A. (Eds.). (1997). *Change and development.* Mahwah, NJ: Erlbaum.

Anderson, T.W. (1960). Some stochastic process models for intelligence test scores. In: K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences*. Stanford, CA: Stanford University Press.

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Bentler, P.M. (1995). *EQS Structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P.M., & Raykov, T. (2000). On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology, 85*, 125-131.

Bijleveld, C.C.J.H., & van der Kamp, L. (Eds.). (1998). *Longitudinal data analysis.* London: Sage.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K.A., & Curran, P.J. (2006). *Latent Curve Models: A Structural Equation Approach*. Hoboken, NJ: Wiley.

Box, G.E.P., Jenkins, P.M. & Reinsel, G.C. (1994). *Time series analysis: forecasting and control*. Englewood Cliffs, NJ: Prentice Hall.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Card, N.A. & Little, T.D. (2007). Longitudinal modeling of developmental processes. *International Journal of Behavioral Development, 31*, 297-302.

Cole, D.A., & Maxwell, S.E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modelling. *Journal of Abnormal Psychology, 112*, 558-577.

Collins, L.M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design and statistical model. *Annual Review of Psychology, 57*, 505-528.

Collins, L.M. & Horn, J.L. (Eds.). (1991). *Best methods for the analysis of change.* Washington, DC: American Psychological Association.

Collins, L.M. & Sayer, A.G. (Eds.). (2001). *New methods for the analysis of change.* Washington, DC: American Psychological Association.

Duncan, T.E., Duncan, S.C., & Strycker, L.A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum.

Dunn, G., Everitt, B. & Pickles, A. (1993). *Modelling covariances and latent variables using EQS*. London: Chapman & Hall.

Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit: ¶

Dwyer, M., Feinlieb, P.L., & Hoffmeister, H. (Eds.) (1992). *Statistical models for longitudinal studies of health.* New York: Oxford.

Embretson, S.E. (2007). Impact of measurement scale in modeling developmental processes and ecological factors. In: T.D. Little, J.A. Bovaird, & N.A. Card (Eds.), *Modeling contextual effects in longitudinal studies*. Mahwah, NJ: Lawrence Erlbaum.

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

Ferrer, E., Balluerka, N., & Widaman, K.F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology, 4,* 22-36.

Fitzmaurice, G.M., Laird, N.M. & Ware, J.H. (2004). *Applied Longitudinal Analysis*. New York: John Wiley.

Frees, E.W. (2004). *Longitudinal and panel data: analysis and applications in the social sciences*. Cambridge: Cambridge University Press.

Goldstein, H., Healy, M.J.R. & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, *13,* 1643-1655.

Gottman, J.M. (Ed.). (1995). *The analysis of change*. Mahwah, NJ: Erlbaum.

Guttman, L.A. (1954). A new approach to factor analysis. The radex. In: P.F. Lazarsfeld (Ed.), *Mathematical thinking the social sciences.* New York: Columbia University Press.

Hayduk, L.A. (1996). *LISREL issues, debates, and strategies*. Baltimore: John Hopkins.

Heise, D.R. (1969). Separating reliability and stability in test re-test correlation. *American Sociological Review, 34,* 93-101.

Humphreys, L.J. (1960). Investigations of the simplex. *Psychometrika, 25*, 313-323.

Jaccard, J., Turrisi, R. & Wan, C.K. (2003). *Interaction Effects in Multiple Regression*. California: Sage Publications.

Jaccard, J., Turrisi, R. & Wan, C.K. (1999). *LISREL approaches to interaction effects in multiple regression released.* California: Sage Pub.

Jöreskog, K.G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrix. In: Atkinson, R.C., Krantz, D.H., Luce, R.D. & Suppes, P. (Eds.), *Contemporary developments in mathematical psychology*. Vol. II. *Measurement, psychophysics, and neural information processing*. San Francisco, CA: Freeman.

Jöreskog, K.G. (1979). A general approach to confirmatory maximum likelihood factor analysis with addendum. In: K.G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Jöreskog, K.G. (1977). Analysis of longitudinal data. In: D.V. Aigner & A.S. Goldelberg (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Jöreskog, K.G. & Sörbom, D. (2001). *LISREL 8 user's reference guide.* Mooresville, IN: Scientific Software.

Jöreskog, K.G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In: G.A. Marcoulides & R.E. Schumacker

Suprimit:
Suprimit: In
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit: ,
Suprimit: In
Suprimit:
Suprimit: *in*
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit:
Suprimit: In
Suprimit: In
Suprimit: In
Suprimit:
Suprimit: In
Suprimit: ¶

(Eds), *Advanced structural equation modeling: Issues and techniques.* Mahwah, NJ: Lawrence Erlbaum.

Judd, C.M. & Kenny, D.A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602-619.

Keesling, J.W. (1972). *Maximum likelihood approaches to causal analysis*. Ph.D. thesis, University of Chicago.

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

Little, T.D., Bovaird, J.A., & Slegers, D.W. (2006). Methods for the analysis of change. In: T.D. Little & D. Mroczek (Eds.), *Handbook of personality development*. Mahwah, NJ: Lawrence Erlbaum.

Little, T.D.; Bovaird, J.A., & Widaman, K.F. (2006). On the merits of orthogonalizing powered and product terms: implications for modeling interactions among latent variables. *Structural Equation Modeling, 13*, 497-519.

Little, T.D., Preacher, K.J., Selig, J.P., & Card, N.A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31*, 357-365.

Little, T.D., Schnabel, K.U. & Baumert, J. (Eds.) (2000). *Modeling longitudinal and multilevel data, practical issues, applied approaches, and specific examples*. Mahwah, NJ: Lawrence Erlbaum.

Little, T.D., Slegers, D.W., & Card, N.A. (2006). A non-arbitrary method identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling, 13*, 59-72.

Marcoulides, G.A. & Drezner, Z. (2001). Specification searches in structural equation modeling with a genetic algorithm. In: G.A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.

McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural analyses. In: R. Cudeck, S. du Toit & D. Sorbom (Eds.), *Structural Equation Modeling: Present and future.* Lincolnwood, IL: Scientific Software International.

McArdle, J.J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110-133.

McArdle, J.J., & Nesselroade, J.R. (2003). Growth curve analysis in contemporary psychological research. In: J.A. Schinka & W.F. Velicer (Eds.), *Handbook of psychology. Research methods in psychology.* Hoboken, NJ: Wiley.

Meredith, W., & Tisak, J. (1984). "Tuckerizing" curves. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In: Collins, L.M. & Sayer, A. (Eds.), *New Methods for the Analysis of Change.* Washington, D.C.: American Psychological Association.

Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463-469.

Suprimit: Stimating

Suprimit:

Suprimit:

Suprimit:

Suprimit: In

Suprimit:

Suprimit:

Suprimit:

Suprimit: Little, T. D., Slegers, D.W., & Card, N.A. (2006). A non-arbitrary method identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling, 13*, 59-72.¶

Suprimit:

Suprimit:

Suprimit: In

Suprimit: In

Suprimit:

Suprimit:

Suprimit: In

Suprimit: In

Suprimit: ¶

Nagin, D.S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4*, 139-157.

Nagin, D.S. (2005). *Group-based modeling of development*. Cambridge: Harvard University Press.

Pearson, K. (1896). Mathematical contributions to the theory of evolution: III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society*, *187*, 253-318.

Ping, R.A. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research, 32*, 336-347.

Plewis, I, (1985). *Analysing change: measurement and explanation using longitudinal data*. Chichester: Wiley.

Plewis, I. (1996). Statistical methods for understanding cognitive growth: A review, a synthesis, and an application. *British Journal of Mathematical and Statistical Psychology, 49*, 25-42.

Plewis I (2005) Modelling behaviour with multivariate multilevel growth curves. *Methodology, 1*, 71-80

Plewis, I, Vitaro, F. & Tremblay, R. (2006). Modelling repeated ordinal reports from multiple informants. *Statistical Modellin, 6*, 251–263.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.

Ram, N. & Grimm, K.J. (2007). Using simple and complex growth to articulate developmental change: Matching theory to method. *International Journal of Behavioral development, 31*, 303-316.

Rindskopf, D. (1984). Using phantom and imaginary variables latent variables to parametrize constraints in linear structural models. *Psychometrika, 49*, 37-47.

Rosel, J. Arnau, J. & Jara, P. (1998). Relationship between the mean and the constant in a Box-Jenkins time-series model. *Quality and Quantity, 32*, 155-163.

Rosel, J. & Elósegui, E. (1991). Un modelo de ecuaciones estructurales para diseños en panel. *Psicothema, 3*, 433-442.

Rosel, J., Jara, P. & Arnau, J. (2002). Geometrical interpretation of the mean and the constant in a Box-Jenkins time-series model. *Quality and Quantity, 36*, 411-425.

Schumacker, R.E. (2002). Latent variable interaction modeling. *Structural Equation Modeling, 9*, 40-54.

Schumacker, R.E. & Marcoulides, G.A., (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schumacker, R.E. & Lomax, R.G. (2004). *A beginner's guide to structural equation modelling*. Mahwah, NJ: Lawrence Erlbaum.

Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6*, 461-464.

Seber, G.A.F. & Wild, C.J. (1989). *Nonlinear regression*. New York: Wiley.

Singer, J.D. & Willet, J.B. (2003). *Applied longitudinal data analysis. Modelling change and event occurrence*. New York: Oxford University Press.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201-293.

Suprimit:

Suprimit:

Suprimit:

Suprimit:

Suprimit: ¶

Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology, 1*, 39-54.

Steyer, R., Eid, M. & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research-Online, 2*, 21-33.

Steyer, R., Krambeer, S. & Hannöver, W. (2004). Modeling Latent Trait-Change. In: K. Van Montfort, H. Oud & A. Satorra (Eds.), *Recent developments on structural equation modeling: theory and applications*. Amsterdam: Kluwer.

Tukey, J.W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

von Eye, A. & Clogg, C.C. (Eds.) (1994). *Latent variables analysis: Applications for developmental research.* Thousand Oaks, CA: Sage.

Wiley, D.E. (1973). The identification problem for structural equation models with unmeasured variables. In: A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar

Wright, S. (1918). On the nature of size factors. *Genetics, 3*, 367-374.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 10*, 557-585.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association, 57*, 348-368.

Suprimit: (www.mpr-online.de).

Suprimit: In

Suprimit:

Suprimit: ¶

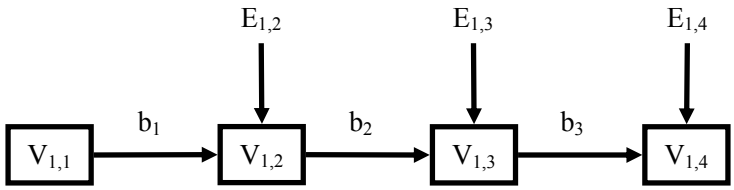


Figure 1. First-order autoregressive univariate model (*AR(1)*), 4W1V.

Suprimit: 1

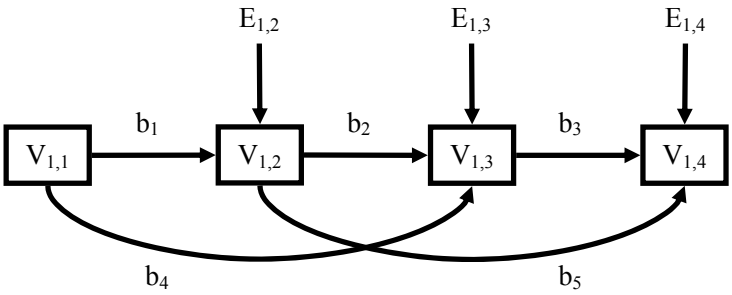


Figure 2. Second-order autoregressive univariate model (*AR(2)*), 4W1V.

Suprimit: 1

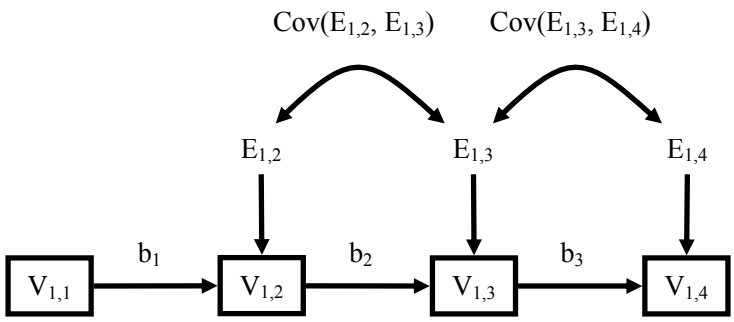


Figure 3. First-order autoregressive model with covariance between the measurement errors (4W1V).

Suprimit: 1

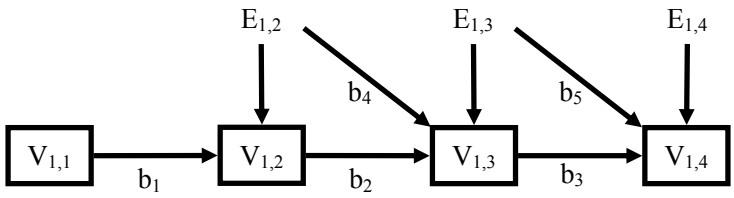


Figure 4. First-order moving average and autoregressive model, *ARMA(1,1),* design 4W1V.

Suprimit: 1

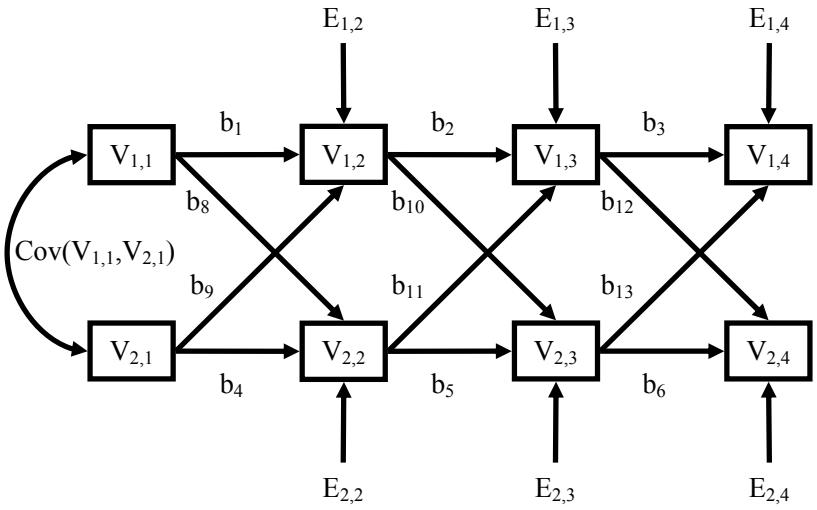


Figure 5. AR(1) bivariate model with crossed effects between the two variables, corresponding to the design 4W2V.

Suprimit: 1

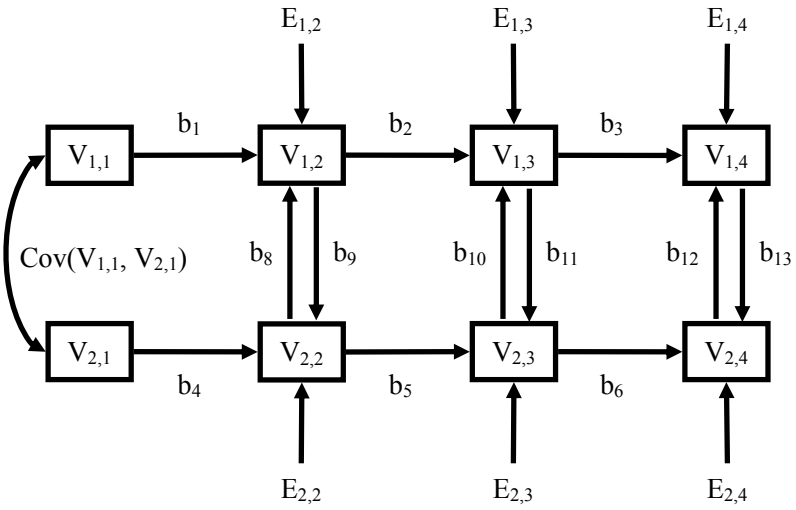


Figure 6. Recursive model with reciprocal influences between variables at the same moment of measurement.

Suprimit: medition

Suprimit: 1

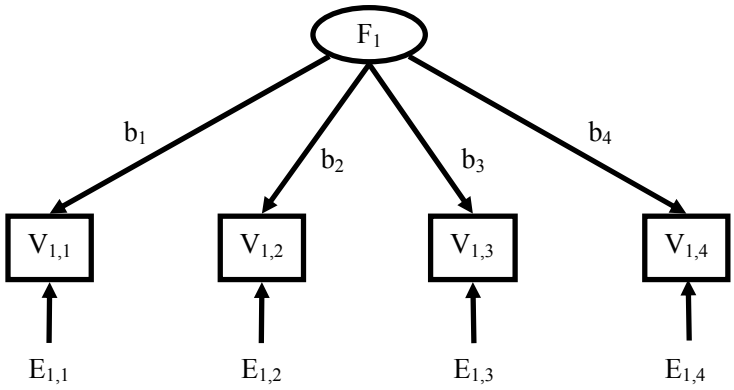


Figure 7. Simple factor model in repeated measures.

Suprimit: 1

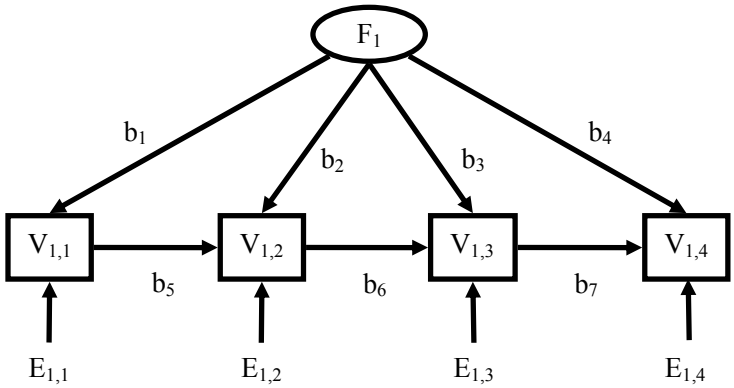


Figure 8. Simple factor model in repeated measures, with AR(1) effects.

Suprimit: 1

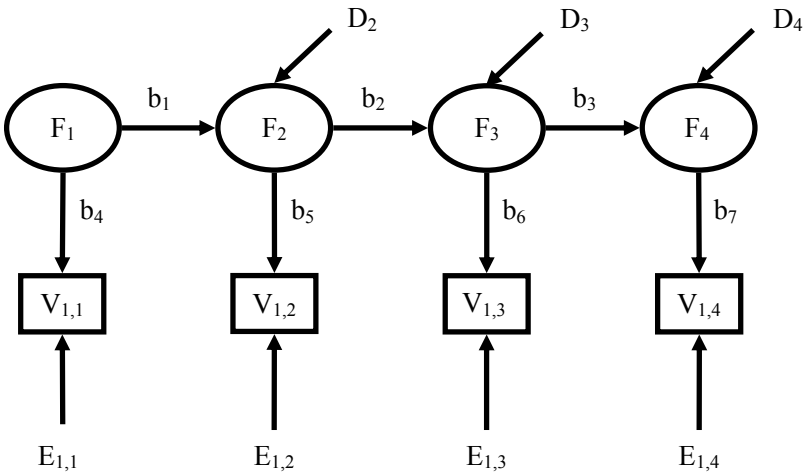


Figure 9. AR(1) univariate model in the factor.

Suprimit: 1

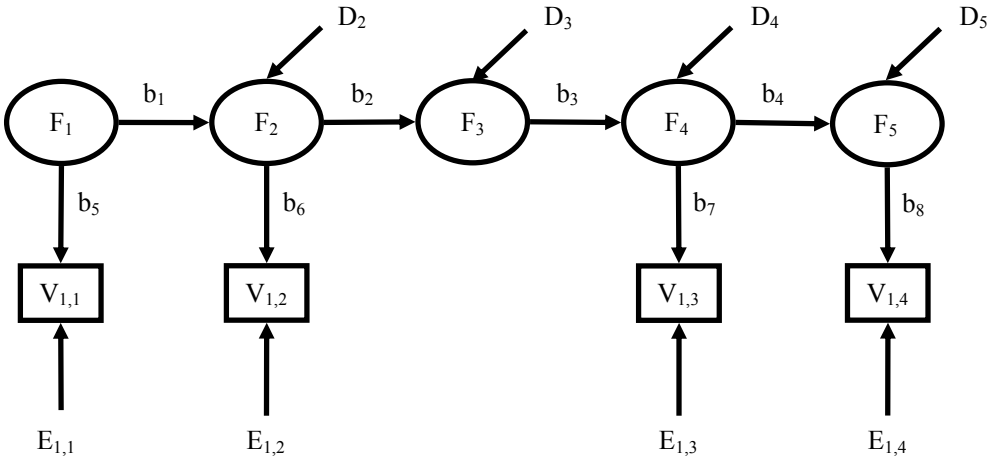


Figure 10. AR(1) univariate model in the factor, with no measure at the time corresponding to F_3.

Suprimit: 1

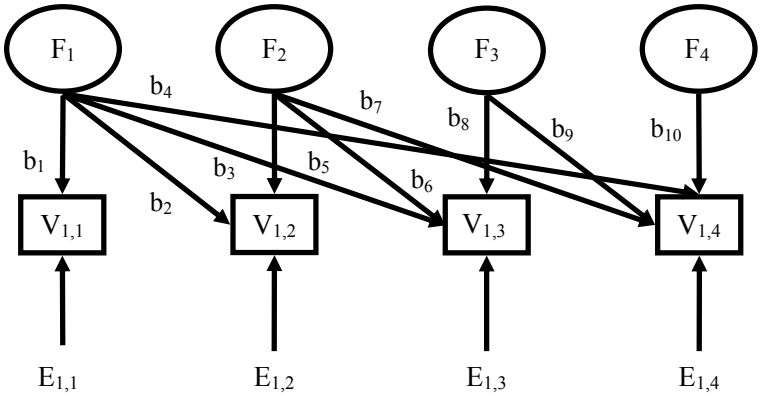


Figure 11. 'Random walk' model of measurement factors.

Suprimit: 1

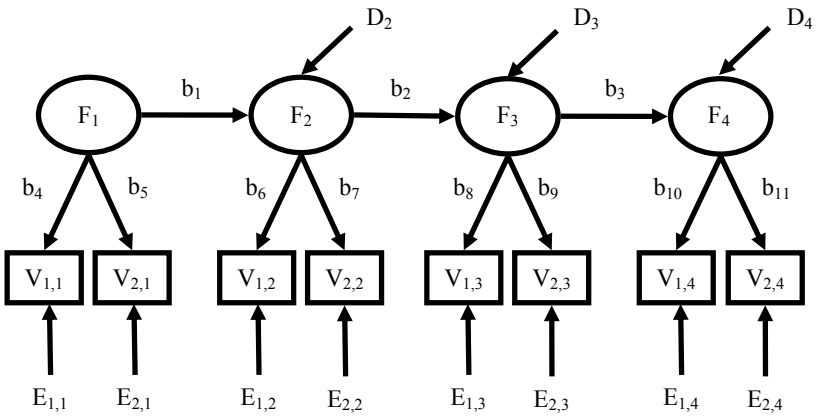


Figure 12. AR(1) bivariate model in the time measurement factors.

Suprimit: 1

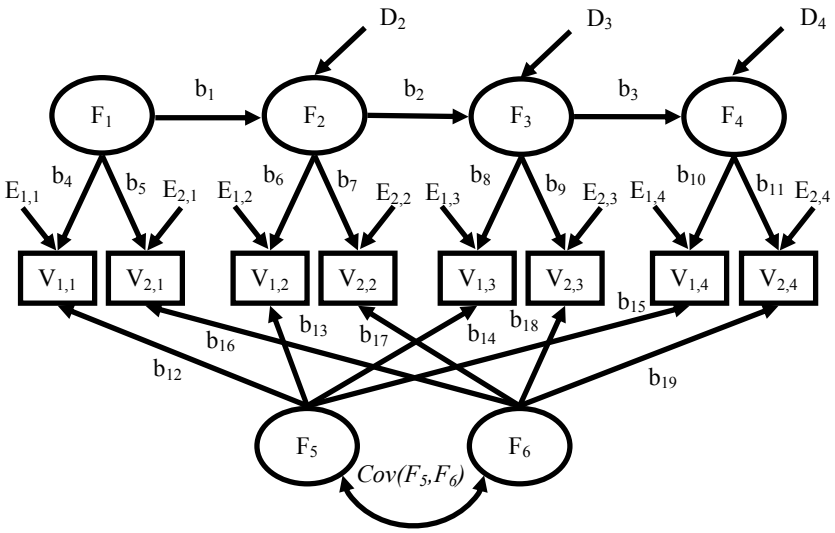
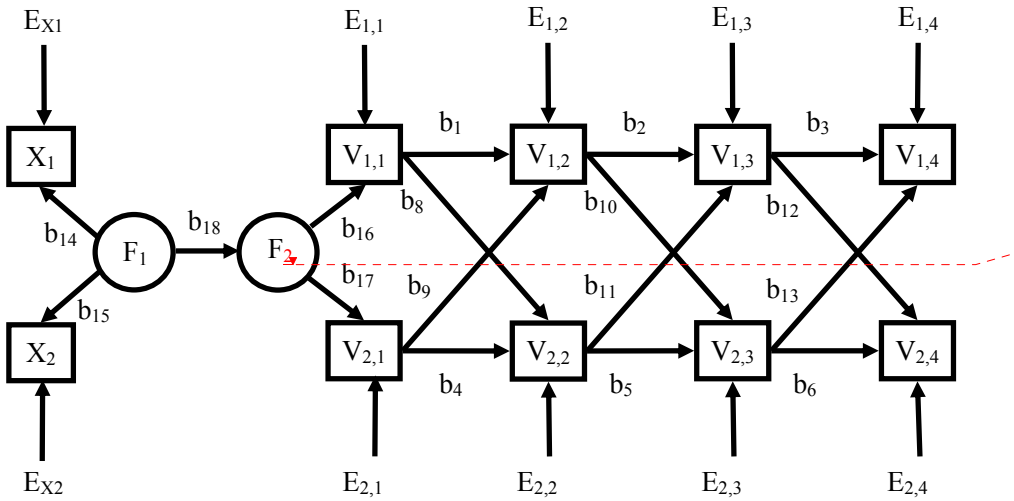


Figure 13. AR(1) bivariate model in time measurement factors (from F_1 to F_4), and with factors indicating the variables (F_5 and F_6).

Suprimit: 1



Suprimit: 1

Figure 14. AR(1) bivariate model with crossed effects between the two variables, with one exogenous latent variable (F_1) that has two observable variables (X_1 and X_2).

Suprimit: 1

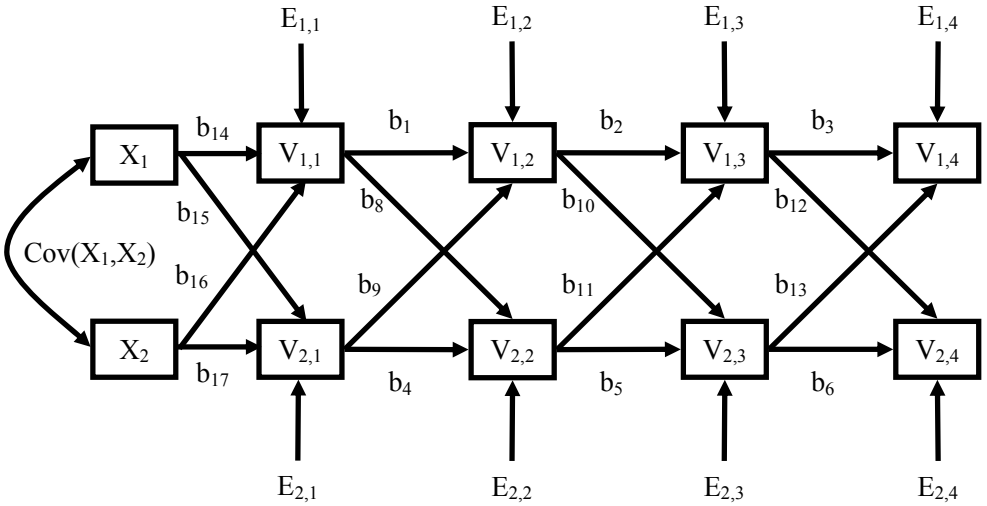


Figure 15. AR(1) bivariate model with crossed effects between the two variables, with two exogenous observable variables (X_1 and X_2).

Suprimit: 1

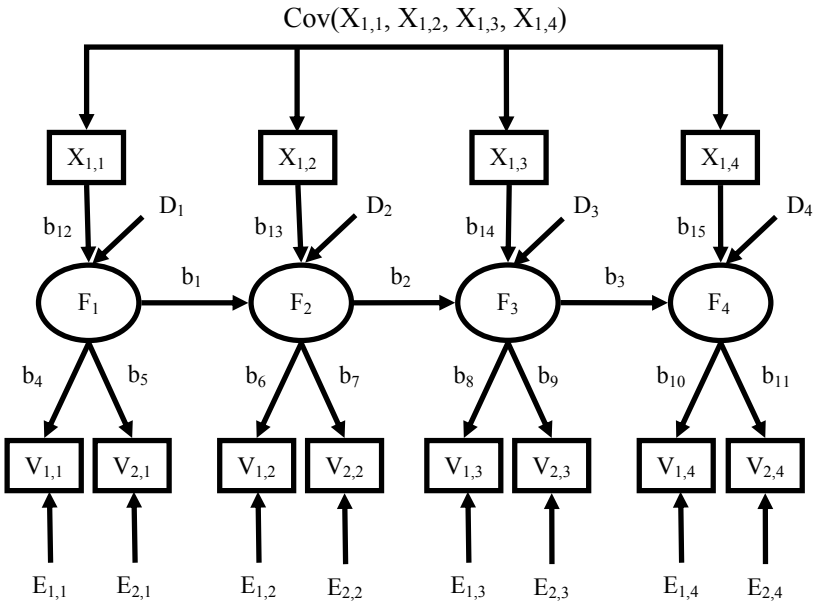


Figure 16. Time-varying model conditioned to the variable $X_{l,t}$. $Cov(X_{1,1}, X_{1,2}, X_{1,3}, X_{1,4})$ indicates all the possible covariances between the variables $X_{1,1}$, $X_{1,2}$, $X_{1,3}$ and $X_{1,4}$: $Cov(X_{1,1}, X_{1,2})$, $Cov(X_{1,1}, X_{1,3})$, … , $Cov(X_{1,3}, X_{1,4})$.

Suprimit: 1

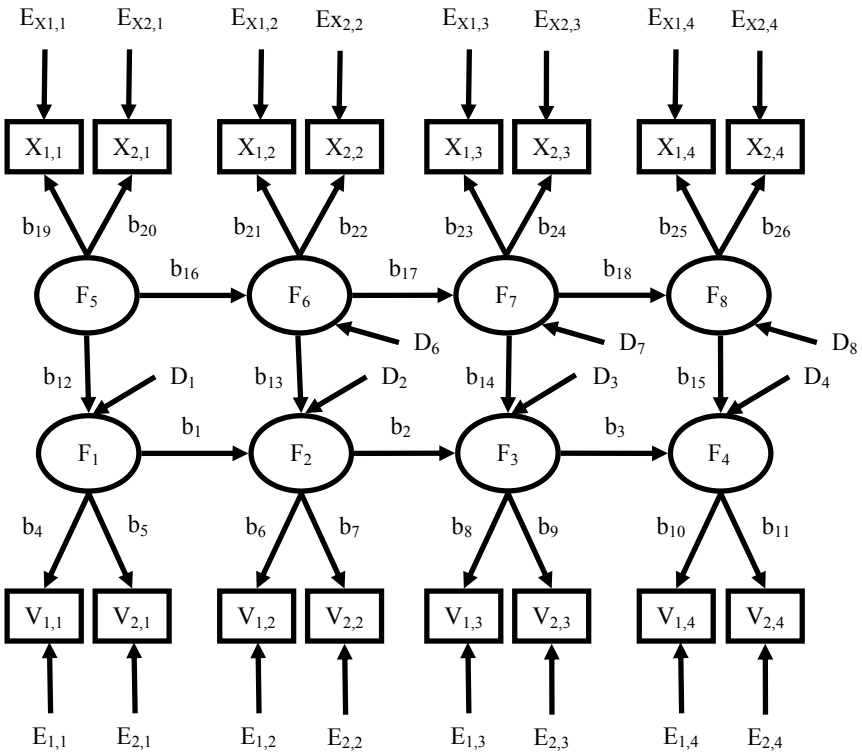


Figure 17. Time-varying model conditioned to the intellectual capacity time factor, indicated from F_5 to F_8.

Suprimit: 1

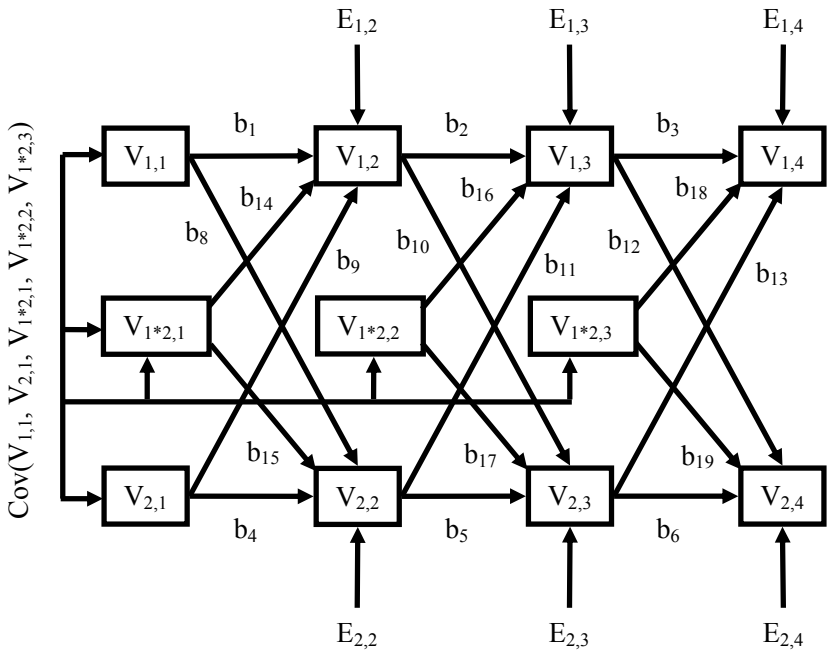


Figure 18. Variable interaction model, with effects on the next time of measurement. Cov($V_{1,1}$, $V_{2,1}$, $V_{1*2,1}$, $V_{1*2,2}$, $V_{1*2,3}$) indicates all the possible (pairwise) covariances between the variables.

Suprimit: 1

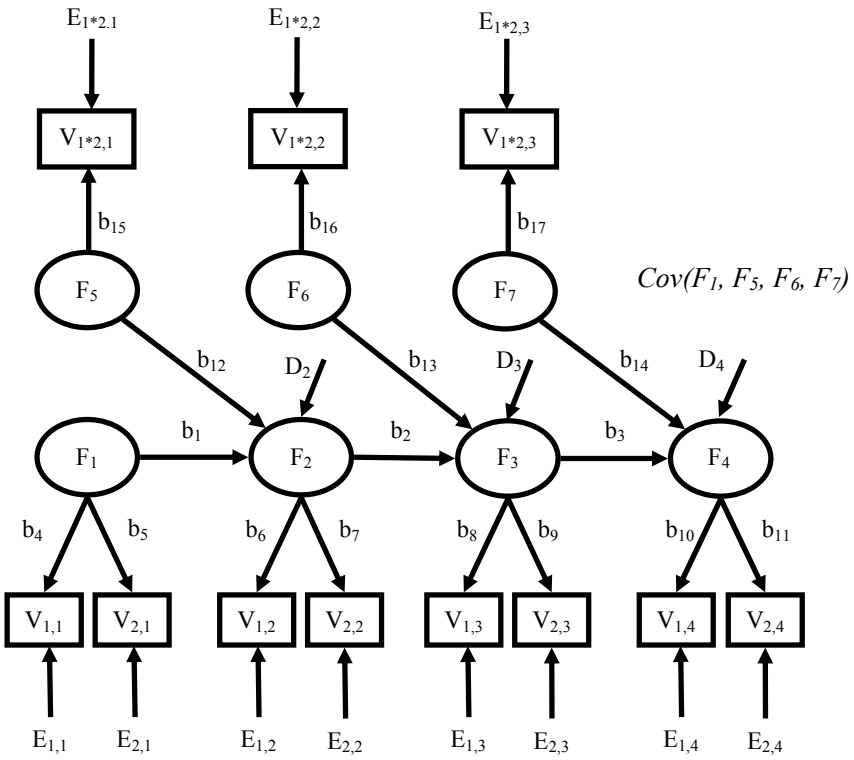


Figure 19. Observed variable interaction model within the measurement time factors. Double-headed arrows must be added to the model to represent the covariances between the independent variables: *Cov(F_1, F_5, F_6, F_7)= Cov(F_1, F_5), Cov(F_1, F_6), …, Cov(F_6, F_7)*.

Suprimit: .

Suprimit: 1

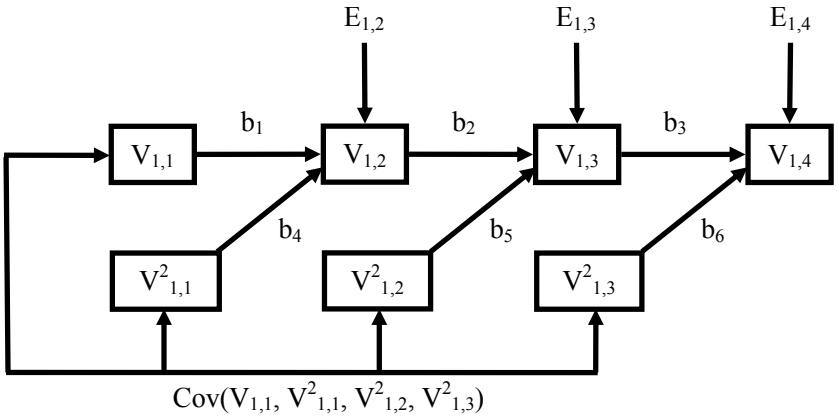


Figure 20. Lag 1 quadratic effects model. Cov($V_{1,1}$, $V^2_{1,1}$, $V^2_{1,2}$, $V^2_{1,3}$) indicates all the (pairwise) covariances of the independent variables: Cov($V_{1,1}$,$V^2_{1,1}$), Cov($V_{1,1}$,$V^2_{1,2}$), …, Cov($V^2_{1,2}$,$V^2_{1,3}$).

Suprimit: 1

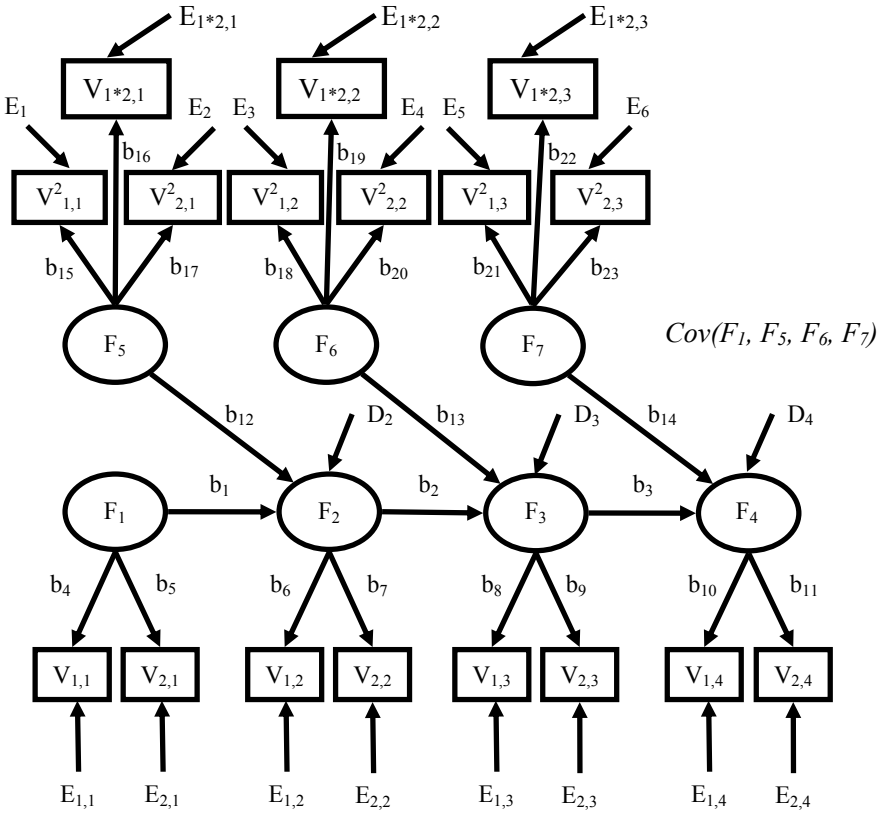


Figure 21. Model of quadratic effects of one factor (F^2_t) on another factor (F_{t+1}). Double-headed arrows must be added to the model to represent the covariances between the independent variables: $Cov(F_1, F_5, F_6, F_7) = Cov(F_1, F_5), Cov(F_1, F_6), \ldots, Cov(F_6, F_7)$.

Suprimit: 1

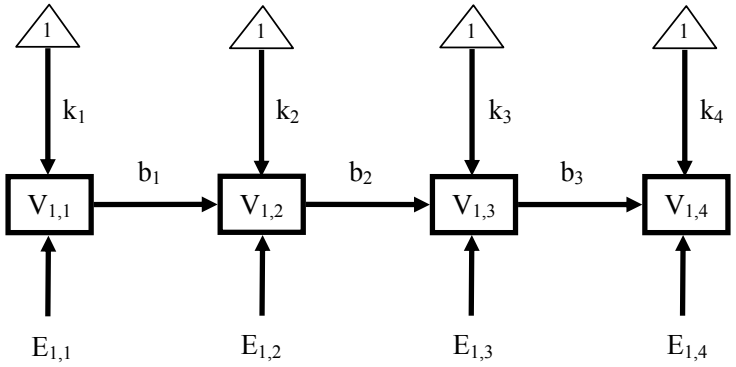


Figure 22. Univariate model *AR(1)* with constants in the variables.

Suprimit: 1

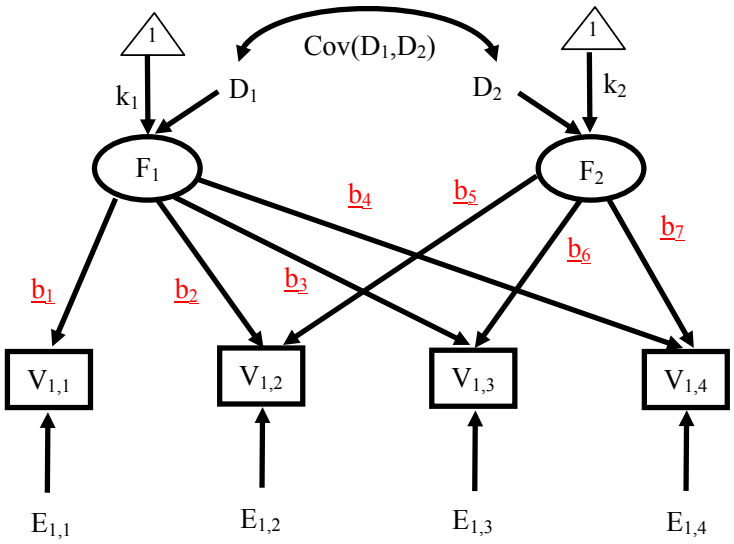


Figure 23. Linear latent growth curve model, with a random constant and a random slope.

Suprimit: .

Suprimit: 1

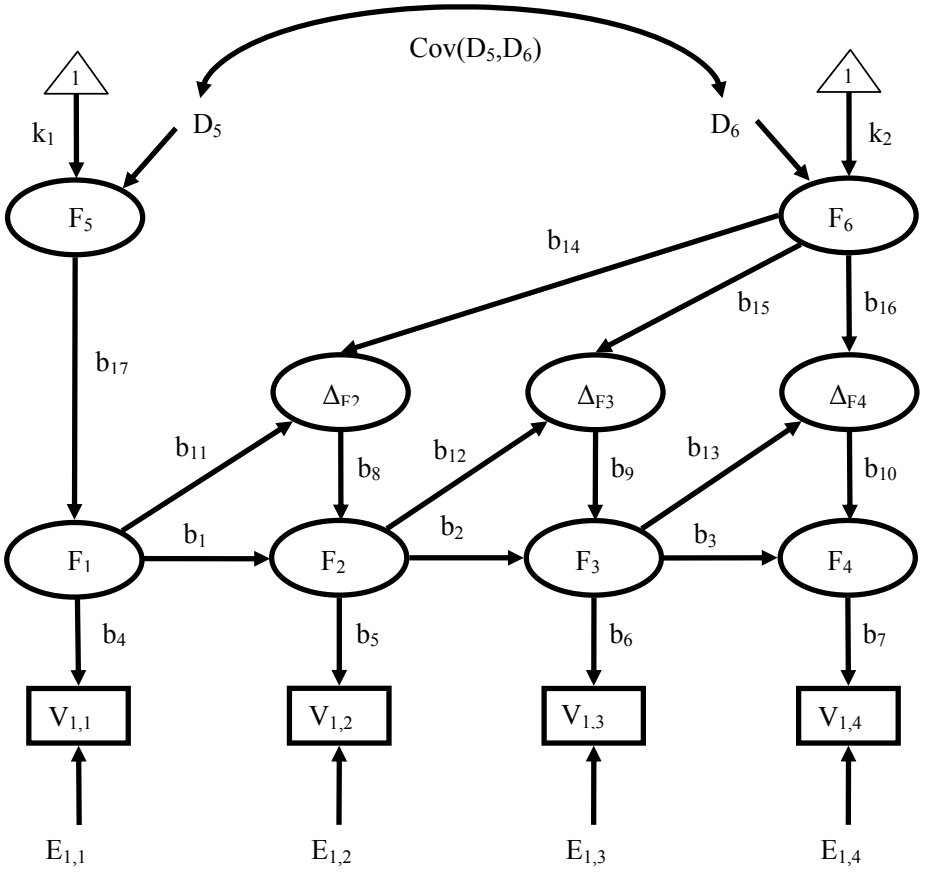


Figure 24. Latent growth difference score model, with a constant initial level and an additive coefficient.

Suprimit: random

Suprimit: random

Suprimit: 1