Noname manuscript No.
(will be inserted by the editor)

Simultaneous EM estimation of a Structural Equation Model and its latent factors

M. TAMI · X. BRY · C. LAVERGNE

Received: date / Accepted: date

Abstract In this work, an EM estimation algorithm of a Structural Equation Model (SEM) and its Latent Variables (LVs) is proposed. Unlike the more prominent Covariance-Based SEM (CBSEM) approach, this estimation is not based on the constrained estimation of the covariance structure of the data. Latent variables are considered as missing data and the EM algorithm is used to maximize the likelihood of the entire model, providing simultaneously estimators of the model's coefficients and predictions of LVs. Contrary to CBSEM which does not take into consideration the structural part (equations of LVs exclusively) of the SEM for the LVs prediction, this EM approach considers the whole data and the complete model. In such context, when LVs are latent factors, SEM includes factorial models in add to a structural part. Then contrary to the EM algorithm for maximum likelihood factor analysis, this work extends the EM estimation in order to take into account the structural part of the SEM (links between the latent factors). Through a simulation study, accuracy and algorithmic performances are investigated. The prevail approaches CBSEM and PLS-PM are compared to the EM estimation according to different criteria. Finally, this approach is applied to a real environmental dataset, providing interesting conclusions.

Keywords structural equation model · latent factors · EM algorithm · maximum likelihood · latent variables prediction

Myriam TAMI
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
Tel.: +336-14-627953
E-mail: myriam.tami@univ-grenoble-alpes.fr

Xavier BRY
University of Montpellier - Institut Montpelliérain Alexander Grothendieck (IMAG), CNRS, Montpellier, France

Christian LAVERGNE
University Paul-Valéry Montpellier 3 - Institut Montpelliérain Alexander Grothendieck (IMAG), CNRS, Montpellier, France

Mathematics Subject Classification (2000) 62Fxx · 62Hxx · 90-08 · 62Jxx

1 Introduction

When it comes to modeling phenomena involving indirect measurements, SEMs are handy and, as such, widely used. These allow to formalize the dependence links of continuous observed variables (Bollen, 1989) through the use of fewer unobserved ones. Every LV is assumed to be underlying a specific set of OVs and summarizes their information. The set of these links are formalized by a system of mathematical equations: measurement equations (describing the links between the LVs and OVs) and structural equations (describing the links between the LVs only). Such models originate in three different fields: path analysis (Wright, 1921; Duncan, 1966), simultaneous-equation models (Haavelmo, 1943; Koopmans, 1945), and factor analysis (Spearman, 1904; Lawley, 1940; Anderson and Rubin, 1956), for which LVs are then considered as unknown unconstrained factors having a known distribution (typically standard normal). Several multivariate statistical techniques were proposed to handle the estimation of these models. Historically, in the early 1970s, these three fields were merged; even though many researchers have significantly contributed (Jöreskog, 1970; Hauser and Goldberger, 1971; Zellner, 1970; Keesling, 2016; Wiley et al, 1973; Browne, 1974), it was the contributions of K. Jőreskog on covariance-based SEM (CBSEM) Jöreskog (1973), Jöreskog (1978), that prevailed the field. The CBSEM approach estimates SEM parameters so that the discrepancy between the estimated and sample covariance matrices is minimized. CBSEM (such as factor analysis) requires distributional assumptions; then, a constrained maximum likelihood estimation is proposed for this variance-covariance matrix. This approach was extended by K. Jőreskog to a more general SEM which involves LInear Structural RELations (LISREL) between LVs (Jöreskog, 1970; Jöreskog and Sörbom, 1982). Indeed, with D. Sőrbom, they developed the computer program LISREL, providing many applied researchers access to the field of SEMing. LISREL was distributed commercially and few years later several new commercial software packages for SEMing were developed, such as EQS (Bentler, 2006), AMOS (Arbuckle, 2011) and Mplus (Muthén and Muthén, 2010). For example, works of Muthén addressed several issues, such as General Linear Modeling, missing data in OVs by means of the EM algorithm (Muthén et al, 1987; Muthén and Muthén, 1998). These commercial tools are still used even if some free computer programs are currently available such as lavaan R package or gllamm which is developed in Stata (Rabe-Hesketh et al, 2004).

CBSEM approach is theoretically well grounded, but has some drawbacks. First, the direct maximization of the likelihood of the variance-covariance matrix is technically hard, and all the more so as the structural equation system is complex. Secondly, concentrating on the variance-covariance matrix, this approach does not *directly* provide estimates of the factor-values at unit

level (called individual scores). Indeed, Unweighted Least Squares (ULS) has been proposed by McDonald (1996), that adds to Jőreskog's assumptions the constraint that LVs are linear combinations of the observed ones and later, Jöreskog (2000) completed the CBSEM approach with a second step to estimate LVs scores. However, this method is based on a least squares technique performed on the mere measurement equations, *neglecting the structural equations* of the SEM.
To overcome such drawbacks, an alternative approach to SEM has been proposed by Wold (1966, 1973, 1982, 1985) and Lohmöller (1989): the Partial Least Squares Path-Modeling (PLS-PM). LVs are then considered as components, i.e. LVs scores are estimated as exact linear combinations of their associated OVs and handles them as error free substitutes for the OVs. Contrary to CBSEM approach, no assumption is made as to their distribution, and only least-square technique is involved in the estimation. This point can be useful for situations where the data is not normally distributed. One should note that PLS-PM is not to be confused with PLS regression (Hair et al, 2011). In PLS-PM the explained variance of the endogenous LVs is maximized by estimating partial model relationships in an iterative sequence of ordinary least squares. It has been shown that this method is computationally efficient and more robust at convergence. Moreover, the components can be predicted from OVs and so, scores are a direct output of this approach. Another advantage of this approach is the ability to predict the values of dependent variables for new statistical units. In spite of such commodities, the PLS-PM approach does not deal appropriately with the partial relationships between components involved in a multiple regression equation, as shown in (Bry and Verron, 2015), who lately proposed an extended method to extract components suiting a multiple equation model: THEME (Bry et al, 2012; Bry and Verron, 2015). Now, constraining LVs to be components may be considered unnecessarily limiting when one is not only interested in prediction, and the factoring approach still makes more sense in this respect.
The present work proposes to focus on tackling this issue by viewing the factors' values as missing data and using the EM algorithm (initially designed by Dempster et al (1977)) to maximize the likelihood of the whole data. Extending the use of the EM algorithm in a related way was proposed in the framework of mixed linear models by Dempster et al (1981); Andrade and Helms (1984) and then, in the framework of factor analysis (Rubin and Thayer, 1982). Rubin and Thayer (1982) apply the EM algorithm on factorial models that can be compared to a part of a SEM: the measurement equations. The great advantage of EM over the classical Newton-Raphson, Fisher Scoring and Fletcher and Powell algorithms is that the EM algorithm automatically allows to keep parameters in their space and does not require to compute the hessian matrix on each step. However, there exists some well-known potential problems of this EM algorithm such as the ones pointed out by Bentler and Tanaka (1983) on its slow convergence. Moreover, whereas Dempster et al (1977) prove that even if the starting point is one where the likelihood is not convex, if an instance of the algorithm converges, it will converge to a (local) maximum of

the likelihood; this point is contested by Bentler and Tanaka (1983) and led to the problem that EM for multidimensional optimization might tend to stop at a point due to its slow convergence rather than due to a true optimal point being reached. To tackle this issue, the initialization step is very important. An initialization close to the maximum of the likelihood is useful to allow to the EM estimation algorithm to converge to the maximum of the likelihood. This work focuses on this important point and proposes an initialization step based on PCA to address this EM algorithm weakness. Although CBSEM only considers Full-Information Maximum Likelihood (FIML, (Arbuckle et al, 1996)), using the EM algorithm has been considered by Muthén et al (1987); Tang and Lee (1998); Lee and Tang (2006) in order to deal with censoring when it occurs according to a known process. Our approach is different from all previous ones, in that we neither consider the missing data to be partial or censored. Nor do we consider the likelihood of the mere constrained sample variance-covariance matrix, but that of the whole data and the complete model, which is not the case in (Rubin and Thayer, 1982) that applies EM algorithm on factorial equations corresponding to a part of a SEM. The proposed EM estimation approach proposed focuses also on the structural equations (equations of the latent factors only). Then, contrary to (Rubin and Thayer, 1982), the EM algorithm is used to estimate the SEMs' parameters and factors.
To keep the developments simple in the paper, the SEM is restricted to only one structural equation, but it does not lessen the generality of our EM approach. The rest of this paper is organized as follows. Section 2 formally introduces the equations of the SEM we deal with. Section 3 presents the methodological details of the EM algorithm application to the SEM and shows how the EM algorithm estimation provides simultaneously both parameter and latent factor estimates. Section 4 first presents a simulation-based study of the performance of this EM approach, with comparison to prevail methods: CBSEM and PLS-PM, and then illustrates an application to environmental data.

2 The SEM specification and notations

The data consists in blocks of OVs describing the same n units. To improve readability, we use the notation A' and v' to specify the transpose of a matrix A and a column vector v, respectively. Let, $X^m = \{x^m_{i,j}\}$; $i \in \{1, \ldots, n\}$ (the row index), $j \in \{1, \ldots, q_m\}$ (the column index), be the $n \times q_m$ matrix coding the explanatory block $m \in \{1, \ldots, p\}$ of OVs with $x_i^{m\prime} = \left(x^m_{i,1}, \ldots, x^m_{i,q_m}\right)$ the q_m-length row vector for the ith observation. Variable blocks refer to the corresponding matrices. Let $Y = \{y_{i,j}\}$; $i \in \{1, \ldots, n\}$ (the row index), $j \in \{1, \ldots, q_Y\}$ (the column index) be the $n \times q_Y$ matrix coding the dependent block of OVs with $y'_i = (y_{i,1}, \ldots, y_{i,q_Y})$ the q_Y-length row vector for the ith observation. Each block of OVs is depending on a factor: Y depending on a factor g and each explanatory block X^m on a factor f^m. T (resp. $T^1, \ldots, T^p$) refers to a $n \times r_T$ (resp. $n \times r_1, \ldots, n \times r_p$) matrices of covariates. We assume n observations; hence the rows of matrices $Y, X^1, \ldots, X^p$ are independent and multivariate normal vectors. The SEM that we handle here consists of $p+2$ equations and a diagrammatical representation is shown in Figure 1.

For each equation in this model, each observed variable in a block is expressed as a linear combination of the corresponding factor, the covariates, and some noises. Hence, the model,

$$\begin{cases} Y & = TD + gb' + \varepsilon^Y \\ \forall m \in \{1,\dots,p\}, X^m & = T^m D^m + f^m a^{m\prime} + \varepsilon^m \\ g & = f^1 c^1 + \cdots + f^p c^p + \varepsilon^g \end{cases} \tag{1}$$

The corresponding equation set of the model (1), for a given observation i and only two explanatories factors, reads

$$\begin{cases} y_i' & = t_i{}' D + g_i b' + \varepsilon_i{}^{y\prime} \\ x_i^{1\prime} & = t_i^{1\prime} D^1 + f_i^1 a^{1\prime} + \varepsilon_i{}^{1\prime} \\ x_i^{2\prime} & = t_i^{2\prime} D^2 + f_i^2 a^{2\prime} + \varepsilon_i{}^{2\prime} \\ g_i & = f_i^1 c^1 + f_i^2 c^2 + \varepsilon_i{}^g \end{cases} \tag{2}$$

where D (resp. D^m) is a $r_T \times q_Y$ (resp. $r_m \times q_m$) parameter matrix, b (resp. a^m) a $1 \times q_Y$ (resp. $1 \times q_m$) parameter vector, c^m a scalar parameter, ε^Y (resp. ε^m) an $n \times q_Y$ (resp. $n \times q_m$) measurement-error matrix. And all the factors are n-length vectors. We impose that the first column of T as well as of each T^m matrix is equal to the constant vector having all elements equal to one. Thus, the first row of D and each D^m contain mean-parameters.
The SEM we handle here is a restricted one, containing only one structural equation, relating a dependent latent factor g (underlying block Y), to p explanatory latent factors $f^1,\dots,f^p$ (underlying respectively blocks $X^1,\dots,X^p$).

The main assumptions of this model are as follows. As far as distributions are concerned, we assume for all $m \in \{1,\dots,p\}$ that $\varepsilon^g \sim \mathcal{N}(0,1)$ (The unit-variance of noise ε^g serves an identification purpose), $\varepsilon_i^Y \sim \mathcal{N}_{q_Y}(0, \psi_Y)$, with $\psi_Y = diag(\sigma^2_{Y,j})_{j\in\{1,\dots,q_Y\}}$, $\varepsilon_i^m \sim \mathcal{N}_{q_m}(0,\psi_m)$, with $\psi_m = diag(\sigma^2_{m,j})_{j\in\{1,\dots,q_m\}}$. Regarding the factors, we assume g is normal with zero-mean, and its expectation conditional on $f^1,\dots,f^p$ is a linear combination of them, then for all $m \in \{1,\dots,p\}$ we assume that $f^m \sim \mathcal{N}_n(0, I_n)$ with $f^1,\dots,f^m$ independent. In each block, we assume that the observed variables $(X^m)_{m\in\{1,\dots,p\}}$ depend linearly on the block's factor $(f^m)_{m\in\{1,\dots,p\}}$ and a block of covariates[1] $(T^m)_{m\in\{1,\dots,p\}}$, conditional on which they are independent. We assume Y depends linearly on the factors $f^1,\dots,f^p$ and an covariate T. Finally, we assume that ε^Y and ε^m, ε^g and f^m are mutually independent for all $m \in \{1,\dots,p\}$.

In order to both avoid heavy formulas in the development of the algorithm and simplify the corresponding implementation, we shall use in the sequel, with no loss of generality, the simplified model (2) involving $p=2$ explanatory blocks X^1 and X^2.

[1] Including covariates allows to remove their effect on the OVs from that of the factor, and thus, look for a better focused factor, which is very important in many practical situations.

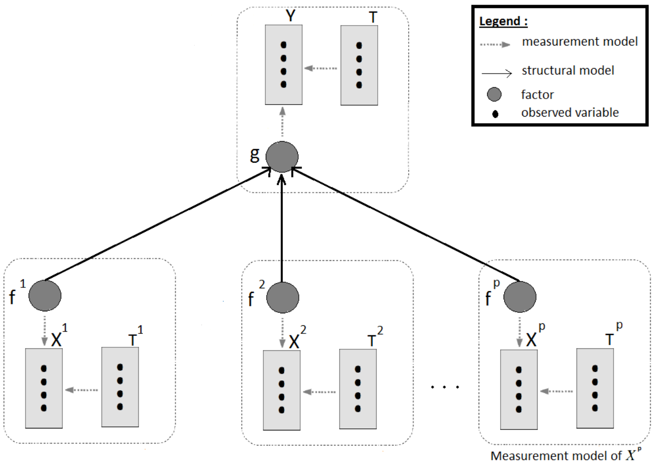

Fig. 1: The Structural Equation Model.

3 Estimation using the EM algorithm

The simultaneous estimation of the SEM's parameters and factors consists to carry out likelihood maximization through an EM algorithm (Dempster et al (1977), Section 4.7). Each iteration of the algorithm involves an Expectation (E)-step followed by a Maximization (M)-step. Although Dempster et al (1977) prove that the EM algorithm yields maximum likelihood estimates[2], some weaknesses in this algorithm are discussed in (Bentler and Tanaka, 1983), answered by Rubin and Thayer (1983). They present the problem that EM for multidimensional optimization might tend to stop at a point due to its slow convergence rather than due to a true optimal point being reached. To tackle this issue, the initialization step is very important. An initialization close to the maximum of the likelihood will be presented in this section which allows to the EM estimation method to converge to this maximum in few iterations. A major advantage of the EM algorithm is that it can be used to predict missing values through their expectation conditional on the observed data. Thus, if we consider LVs as missing data, the EM algorithm is an adequate tool to maximize the likelihood of a statistical model involving LV's, but also to predict these LVs, which is the contribution of this work. In this SEM framework, the LVs correspond to the factors. Thus, EM enables to simultaneously estimate both the factors at unit-level and parameters of a complete SEM. The algorithm is described on the model (2) with no loss of generality on both writing

[2] More precisely, they proved that even if the starting point is one where the likelihood is not convex, if an instance of the algorithm converges, it will converge to a (local) maximum of the likelihood.

of the SEMing and the algorithm development.

Let $z = (y, x^1, x^2)'$ be the OVs and $h = (g, f^1, f^2)'$ the factors corresponding to LVs. The EM algorithm is based on the log-likelihood associated with the complete data (z, h). Let $p(z, h; \theta)$ denote the probability density function of the complete data. Concerning the model (2) the complete log-likelihood function is as follows:

$$\begin{aligned}\mathcal{L}(\theta; z, h) = & -\frac{1}{2}\sum_{i=1}^{n}\{ln|\psi_Y| + ln|\psi_1| + ln|\psi_2| \\ & + (y_i - D't_i - g_i b)'\,\psi_Y^{-1}\,(y_i - D't_i - g_i b) \\ & + \left(x_i^1 - D^{1\prime} t_i^1 - f_i^1 a^1\right)'\psi_1^{-1}\left(x_i^1 - D^{1\prime} t_i^1 - f_i^1 a^1\right) \\ & + \left(x_i^2 - D^{2\prime} t_i^2 - f_i^2 a^2\right)'\psi_2^{-1}\left(x_i^2 - D^{2\prime} t_i^2 - f_i^2 a^2\right) \\ & + \left(g_i - c^1\ f_i^1 - c^2\ f_i^2\right)^2 + \left(f_i^1\right)^2 + \left(f_i^2\right)^2\} + \lambda\end{aligned} \tag{3}$$

where λ is a constant and $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$ is the following K-dimensional set of parameters, with

$$K = 5 + q_Y(r_T + 1) + \sum_{m=1}^{2} q_m(r_m + 1).$$

Indeed, to avoid heavy formulas in the development $\psi_Y = \sigma_Y^2 I_{q_Y}$, $\psi_1 = \sigma_1^2 I_{q_1}$ and $\psi_2 = \sigma_2^2 I_{q_2}$ are assumed.

3.1 Estimation of the SEM and its factors

To estimate the SEM, the function (3) has to be maximized. In the framework of the EM algorithm (Foulley, 2002), the following equation is solved: To maximize this function, in the EM framework (Foulley, 2002), we must solve:

$$\mathbb{E}_z^h\left[\frac{\partial}{\partial\theta}\mathcal{L}(\theta; z, h)\right] = 0. \tag{4}$$

where $\mathbb{E}_z^h\left[\frac{\partial}{\partial\theta}\mathcal{L}(\theta; z, h)\right] := \int \frac{\partial}{\partial\theta}\mathcal{L}(\theta; z, h)\, p_z^h(\theta)\, dh = \int \frac{\partial}{\partial\theta}\mathcal{L}(\theta; z, h)\, p(h|z; \theta)\, dh$, with $p_{z_i}^{h_i}$ is the conditional density function of h_i conditional on z_i for the ith observation. To solve the Equation (4), the derivatives of the log-likelihood function given in Equation (3) are needed. Classically, the use of numeric methods are needed, but we deal with Gaussian distributions in this specific case, and this causes $p_{z_i}^{h_i}$ to be an explicit conditional Gaussian distribution. Let us introduce the following notations:

$$p_{z_i}^{h_i} := \mathcal{N}_3\left(M_i = \begin{pmatrix} m_{1i} \\ m_{2i} \\ m_{3i}\end{pmatrix}, \Sigma = \begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{pmatrix}\right)$$

$$
\begin{aligned}
&\widetilde{g_i} := \mathbb{E}_{z_i}^{h_i}[g_i] := {m_1}_i; \quad \widetilde{\gamma_i} := \mathbb{E}_{z_i}^{h_i}[g_i^2] = (\mathbb{E}_{z_i}^{h_i}[g_i])^2 + \mathbb{V}_{z_i}^{h_i}[g_i] = {m_1}_i^2 + \sigma_{11},\\
&\widetilde{f_i^1} := \mathbb{E}_{z_i}^{h_i}[f_i^1] = {m_2}_i; \quad \widetilde{\phi_i^1} := \mathbb{E}_{z_i}^{h_i}[(f_i^1)^2] = (\mathbb{E}_{z_i}^{h_i}[f_i^1])^2 + \mathbb{V}_{z_i}^{h_i}[f_i^1] = {m_2}_i^2 + \sigma_{22},\\
&\widetilde{f_i^2} := \mathbb{E}_{z_i}^{h_i}[f_i^2] = {m_3}_i; \quad \widetilde{\phi_i^2} := \mathbb{E}_{z_i}^{h_i}[(f_i^2)^2] = (\mathbb{E}_{z_i}^{h_i}[f_i^2])^2 + \mathbb{V}_{z_i}^{h_i}[f_i^2] = {m_3}_i^2 + \sigma_{33}
\end{aligned}
$$

The parameters of the normal distribution $p_{z_i}^{h_i}$ are explicit and their form are described and demonstrated in Appendix B. Expressions of the first-order derivatives of $\mathcal{L}$ with respect to θ are established and described in Appendix C.

3.2 Results

The explicit solution of Equation (4) is then obtained by injecting the first-order derivatives of $\mathcal{L}$ with respect to θ in Equation (4). This procedure is also described in Appendix C and for the model (2), the following system characterizing the solution of (4) is obtained:

$$
\begin{cases}
\hat{b} = \frac{(\overline{\widetilde{g}y} - \overline{yt'})(\overline{tt'})^{-1}\ \overline{\widetilde{g}t}}{\overline{\widetilde{\gamma}} - \overline{\widetilde{g}t'}(\overline{tt'})^{-1}\overline{\widetilde{g}t}}\\
\widehat{a^m} = \frac{\overline{\widetilde{f^m}x^m} - \overline{x^m t^{m\prime}}(\overline{t^m t^{m\prime}})^{-1}\overline{\widetilde{f^m}t^m}}{\overline{\widetilde{\phi^m}} - \overline{\widetilde{f^m}t^{m\prime}}(\overline{t^m t^{m\prime}})^{-1}\overline{\widetilde{f^m}t^m}}\\
\widehat{c^1} = \frac{(\sigma_{12} + \overline{\widetilde{f^1}\widetilde{g}})\overline{\widetilde{\phi^2}} - (\sigma_{13} + \overline{\widetilde{f^2}\widetilde{g}})(\sigma_{23} + \overline{\widetilde{f^1}\widetilde{f^2}})}{\overline{\widetilde{\phi^1}}\,\overline{\widetilde{\phi^2}} - (\sigma_{23} + \overline{\widetilde{f^1}\widetilde{f^2}})^2}\\
\widehat{c^2} = \frac{(\sigma_{13} + \overline{\widetilde{f^2}\widetilde{g}})\overline{\widetilde{\phi^1}} - (\sigma_{12} + \overline{\widetilde{f^1}\widetilde{g}})(\sigma_{23} + \overline{\widetilde{f^1}\widetilde{f^2}})}{\overline{\widetilde{\phi^1}}\,\overline{\widetilde{\phi^2}} - (\sigma_{23} + \overline{\widetilde{f^1}\widetilde{f^2}})^2}\\
\widehat{D'} = (\overline{yt'} - \hat{b}\ \overline{\widetilde{g}t'})(\overline{tt'})^{-1}\\
\widehat{D^{m\prime}} = (\overline{x^m t^{m\prime}} - \widehat{a^m}\,\overline{\widetilde{f^m}t^{m\prime}})(\overline{t^m t^{m\prime}})^{-1}\\
\widehat{\sigma_Y^2} = \frac{1}{nq_Y}\sum\limits_{i=1}^{n}\{||y_i - \widehat{D'}t_i||^2 + ||\hat{b}||^2\widetilde{\gamma_i} - 2(y_i - \widehat{D'}t_i)'\hat{b}\widetilde{g_i}\}\\
\widehat{\sigma_m^2} = \frac{1}{nq_m}\sum\limits_{i=1}^{n}\{||x_i^m - \widehat{D^{m\prime}}t_i^m||^2 + ||\widehat{a^m}||^2\widetilde{\phi_i^m} - 2(x_i^m - \widehat{D^{m\prime}}t_i^m)'\widehat{a^m}\,\widetilde{f_i^m}\}
\end{cases}
\tag{5}
$$

where $\overline{x^m} := \frac{1}{n}\sum_{i=1}^{n} x_i^m$, $\forall m \in \{1, \ldots, p\}$ the average on the index i. This kind of system (solution of Equation (4)) can be obtained such as described by the procedure in Appendix C for any kind writing of SEM (1).

3.3 The algorithm

This algorithm is implemented and available as an R package on GitHub[3]. This EM algorithm estimation for a SEM was applied to a Health-related quality of life longitudinal data (Barbieri et al, 2017; Tami, 2016). Indeed,

[3] https://github.com/myriamtami/EMsem.

after estimations of the latent factors of a SEM produced by this algorithm, a second step is performed to explain the global health status of patients by additional explanatory variables using a linear mixed model.

Input :
– $\{Y, X^1, \ldots, X^p\}$ data matrices: blocks of OVs;
– ϵ, precision for the stopping criterion;

Initialization:
– Initialize the set of the SEM parameters $\theta^{*(0)}$; #θ^* **is the K-dimensional vector containing the values of all scalar parameters in θ.**
– $t \leftarrow 0$;# t **is the iteration index**

$L\_Factors \leftarrow \emptyset$; #**List of predictions of the factors concatenation** $\left(\widetilde{g}, \widetilde{f^1}, \ldots, \widetilde{f^p}\right)$
$L\_Factors\_sq \leftarrow \emptyset$; #**List of the concatenation** $\left(\widetilde{\gamma}, \widetilde{\phi^1}, \ldots, \widetilde{\phi^p}\right)$
repeat
  # **E-step**
  for $i \in \{1, \ldots, n\}$ **do**
    #**Compute the parameters of the distribution** $p_{z_i}^{h_i}$
    $M[i] \leftarrow \mathbb{E}_{z_i}^{h_i}\left[\left(g_i, f_i^1, \ldots, f_i^p\right)\right]$; #**the formula is detailed in Appendix C**
    $\Sigma[i] \leftarrow \mathbb{V}_{z_i}^{h_i}\left[\left(g_i, f_i^1, \ldots, f_i^p\right)\right]$; #**the formula is detailed in Appendix C**
    $M\_sq[i] \leftarrow M[i] \cdot M[i]$;
    $L\_Factors[t] \leftarrow M$; #**prediction of the latent factors**
    $L\_Factors\_sq[t] \leftarrow M\_sq + diag\,(\Sigma)$;
  end
  # **M-step**
  for $k \in \{1, \ldots, K\}$ **do**
    #**Update θ to $\theta^{(t+1)}$ by injecting $\widetilde{g}^{(t)}$, $\widetilde{\gamma}^{(t)}$ and $\widetilde{f^m}^{(t)}$, $\widetilde{\phi^m}^{(t)}$, $m \in \{1, \ldots, p\}$ into Formulas in** (9)
    $\theta^{(t+1)} \leftarrow \hat{\theta}^{(t)}\left(L\_Factors[t], L\_Factors\_sq[t]\right)$;
    $t \leftarrow t + 1$;
  end
until

$$\sum_{k=1}^{K} \frac{|\theta^{*(t+1)}[k] - \theta^{*(t)}[k]|}{|\theta^{*(t+1)}[k]|} < \epsilon$$

Output : Set of the estimations of the parameters $\theta^{(t+1)}$ and the set of the predictions of the factors $L\_Factors[t]$

Algorithm 1: EM algorithm for the simultaneous estimation of a SEM parameters and its latent factors.

In the initialization step, $\forall m \in \{1, \ldots, p\}$ we propose to obtain $D^{m(0)}$ by multiple linear regression of X^m on T^m. Then, to initialize the other parameters, each approximated factor $\widetilde{f^m}^{(0)}$ and $\widetilde{g}^{(0)}$ is computed as first principal component of $X^m - T^m D^{m(0)}$ and $Y - TD^{(0)}$. Then, a^m, σ_m^2 (resp. b, σ_y^2) are initialized through a multiple linear regression of $X^m - T^m D^{m(0)}$ on $\widetilde{f^m}^{(0)}$ (resp. of $Y - TD^{(0)}$ on $\widetilde{g}^{(0)}$). Finally, each $c^{m(0)}$ can be obtained by multiple

linear regression of $\widetilde{g}^{(0)}$ on the p factors $\widetilde{f^m}^{(0)}$. This tricky procedure allows an initialization close to the maximum of the likelihood.
Experiments with the criterion used in the Algorithm (1) has revealed no problems but the user could easily change it as his convenience. A more strict criterion can be used here, for example (Schoenberg and Richtand, 1984) compares the update estimate of each parameter with their corresponding value then, if the absolute value of the previous estimate minus the updated estimate of any of them is greater than some small value ϵ, the algorithm returns to the E-step; otherwise the solution has been reached.

4 Tests and performances of the algorithm

4.1 Numerical results on simulated data

4.1.1 Data generation

To test and evaluate performances of the algorithm, a data generation procedure is performed 100 times, each time yielding a set of simulated data $(Y, X^1, X^2, T, T^1, T^2)$. Each set is composed of $n = 400$ units and $q_Y = q_1 = q_2 = 40$ the number of dimensions in of OVs blocks. Therefore, the 120 OVs Y, X^1, X^2 are simulated so as to be structured respectively around three factors g, f^1, f^2. Factors f^1 and f^2 are explanatory of g. Besides, 2 covariates $r_T = r_1 = r_2 = 2$ are simulated for each covariate matrix T, T^1 and T^2. Each set of simulated data is simulated as follows:

1. Choice of the parameter values θ is as follows:
   (a) $D = D^1 = D^2 =$ are matrices filled row-wise with the ordered integer sequence ranging from 1 to 80 (indeed: $r_T \times q_Y = r_1 \times q_1 = r_2 \times q_2 = 2 \times 40$);
   (b) $b = a^1 = a^2 =$ are ordered integer sequence ranging from 1 to 40;
   (c) $c^1 = c^2 = 1$;
   (d) $\sigma_Y^2 = \sigma_1^2 = \sigma_2^2 = 1$.
2. Simulation of factors g, f^1, f^2 is as follows:
   (a) Simulate $n = 400$ length vectors f^1 and f^2 with a standard normal distribution (abbreviated $\forall m, \in \{1, 2\}$
   $f^m \sim \mathcal{N}_{400}(0, I)$);
   (b) Simulate ε^g according to distribution $\varepsilon^g \sim \mathcal{N}_{400}(0, I)$;
   (c) Calculate $g = f^1c^1 + f^2c^2 + \varepsilon^g$.
3. Simulation of noise matrices ε^Y, ε^1, ε^2:
   Each element of matrix ε^Y, (resp. ε^1, ε^2) is simulated independently from distribution $\mathcal{N}(0, \sigma_Y^2)$ with $\sigma_Y^2 = 1$ as chosen in the first step (resp. $\sigma_1^2 = 1$, $\sigma_2^2 = 1$).
4. Simulation of covariate matrices T, T^1, T^2:
   Each element of matrices T, T^1, T^2 is simulated according to the standard normal distribution.

5. Simulation of Y, X^1, X^2:
   Y, X^1, X^2 are eventually calculated through formulas in the Model (1).

This simulation scheme was performed 100 times, each time yielding a set of simulated data matrices $(Y, X^1, X^2, T, T^1, T^2)$; moreover, an estimation routine with a threshold value $\varepsilon = 10^{-2}$ was also performed, yielding the average results presented in section 4.2. Thus from $400 \times 120 = 48000$ scalar elements of data, $3 \times n = 1200$ scalar elements of factors plus $K = 5 + 3 \times 40(2+1) = 365$ scalar parameters (i.e. 1565 scalars) was estimated.

4.1.2 Results

Convergence is observed in almost all cases in less than five iterations. We assess the quality of the estimations as follows.

- On the one hand, we compute the absolute relative deviation between each simulated scalar parameter in θ and its estimate, and then average these deviations over the 100 simulations. We then produce a box plot of the average absolute relative deviations (see Figure 2). This makes the interpretation easier, since we only need to look at the box plot's values and check that they are positive (because of the absolute value) and close to 0.
- On the other hand, to assess the quality of the factor estimations, we compute the 300 values of square correlations between the simulated concatenated factors (g, f^1, f^2) (respectively) and the corresponding estimates $(\widetilde{g}, \widetilde{f^1}, \widetilde{f^2})$. Once again, we produce a box-plot of these correlations (cf. Figure 3) and check that it indicates values close to 1.

Figures 2 and 3 show clearly that the estimates are very close to the actual quantities. Indeed, on Figure 2, the median of average absolute relative deviations is 0.018, first and third quartiles being 0.015 and 0.023 respectively. On Figure 3, the median of square correlations is 0.998, first and third quartiles being 0.997 and 0.999 respectively. So, factor g (respectively f^1 and f^2) turn out to be drawn towards the principal direction underlying the bundles made up by OVs Y (respectively X^1 and X^2). Now, we may legitimately wonder how the quality of estimations could be affected by the number of observations and the number of OVs in each block. In the following section we provide a sensitivity analysis performed to investigate this issue.

4.1.3 Sensitivity analysis[4] of estimations: empirical behavior study of both the parameter and latent factor estimates

A sensitivity analysis is performed on the sets of simulated data presented in Section 4.1. The purpose is to investigate how the quality of the estimations could be affected by the number n of subjects and the number q_Y,

[4] "Sensitivity analysis" is here used in a different meaning from the "sensitivity analysis" as introduced by Saltelli et al (2000).

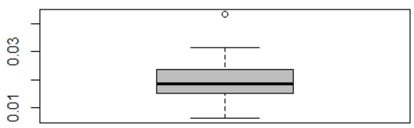

Fig. 2: Box plot of the average absolute relative deviations of the simulated parameters and their estimates.

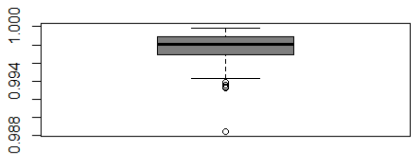

Fig. 3: Box plot of the correlations of the simulated factors and their estimates.

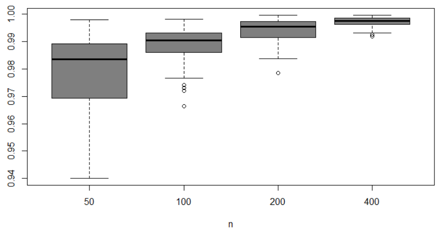

Fig. 4: Box plots of the correlations of simulated factors and their estimates for various values of n.

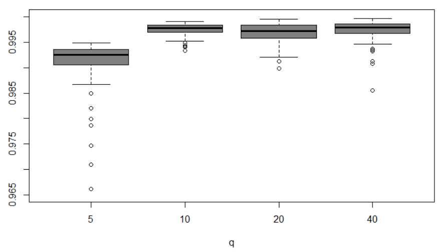

Fig. 5: Box plots of the correlations of simulated factors and their estimates for various values of q.

$(q_m)_{m=1,2}$ of observed variables in each block. To simplify the analysis, we imposed $q_Y = q_1 = q_2 = q$ and varied n and q separately, i.e. studied the cases $n = 50, 100, 200, 400$ with $q = 40$ and $q = 5, 10, 20, 40$ with $n = 400$. Each case was simulated 100 times. Therefore, we simulated 800 data-sets.

Sensitivity with respect to the number of observations n In this section, we study the evolution with n of the average estimation of structural coefficients c^1 and c^2 and parameter σ_Y^2 with respect to their actual values, all equal to 1, and that of the correlations of factors with their estimates. The number of OVs is set to $q = 40$ in each block. Figure 6 graphs these evolutions (average value of estimate in plain line), including an interval constructed by plus and minus the standard deviation (computed on the estimations) about each average estimate (dotted line). This figure shows that the biases and the standard deviations are, as expected, more important for little values of n, but also that the quality of estimation is already quite good for $n = 50$. As for the correlations of factors with their estimates, Figure 4 shows that they increase and get close to one as n increases, with a dispersion decreasing to 0. However, even for $n = 50$, the correlations are mostly above 0.95, indicating that the factors are correctly reconstructed.

Sensitivity with respect to the number q of OVs in each block Likewise, the evolution of the average estimates of c^1, c^2, σ_Y^2 and the correlation of factors with their estimates is studied for different values of q, with n set to 400. Unsurprisingly, the biases and the standard deviations decrease as q

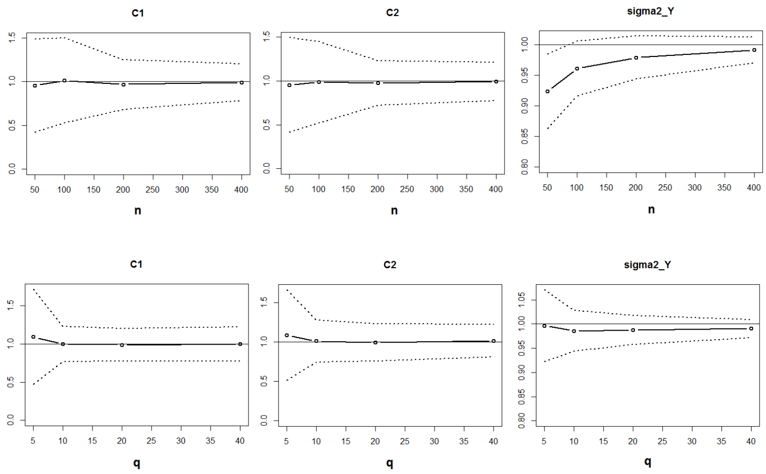

Fig. 6: Average estimates of c^1, c^2, σ_Y^2 and 95% confidence intervals as a function of n and as a function of q.

increases (see Figure 6). In fact, they stabilize even faster with q than with n, particularly σ_Y^2. Indeed, from $q = 10$ on, the confidence interval is narrow enough. As for the factors, Figure 5 shows that their correlations with their estimates are already very close to 1 for $q = 5$, with a very small variance, and keep increasing with q. To sum things up, the sample size n proved to have more impact on the quality of parameter estimation and factor reconstruction than the number of OVs. Now, the quality of factor reconstruction remains high for rather small values of n or q. We advise to use a minimal sample size of $n = 100$ to obtain really stable structural coefficients. Above this threshold, n has but little impact on the biases and standard deviations of estimates.

4.2 Comparison to the prevail methods: CBSEM and PLS-PM

Numerical results obtained in the last section provide useful information to compare the EM estimation approach to the classical methods. Table (1), inspired by (Stan and Saporta, 2006) presents a comparison between CBSEM, PLS-PM and the EM estimation approaches according to different criteria:

Characteristics	CBSEM	PLS-PM	EM-estimation
Goal oriented towards	the SEM's parameters estimation	the SEM's prediction	the SEM's parameters estimation and latent factors prediction
LVs	Factors	Components	Factors
Assumptions	Independent units and multivariate normal distribution	Independent units	Independent units and multivariate normal distribution
Optimality for	parameters estimation accuracy	prediction accuracy	factors prediction and parameters estimation accuracy
Sub-models quality	structural model better because the LVs are estimated in an unrestricted space	measurement model better because LVs are contained in the space of their OVs	trade-off between sub-models because the complete model is used for LVs and parameters estimation
Parameters estimation	**Var-covariance structure and ML estimation**	**Least square technique and regressions**	**ML estimation of the complete model**
LVs estimation or reconstruction	**Least square technique performed on the mere measurement equations**	**Combinations of the OVs they are related to**	**ML estimation of the complete model**
Parameters and LVs estimation of the SEM are handled	separately	simultaneously	simultaneously
Covariables handled by the model	NO	NO	**YES**
Convergence	matrices must to be not singular	observed for more than two blocks	both
Complexity of the model	medium: less than < 100 OVs	great, for example: 100 LVs, 1000 OVs	great: more than 4 OVs by block
Minimal number of units	great ($n \in \{200, \ldots, 800\}$)	few ($n \in \{30, \ldots, 100\}$)	medium ($n > 100$)
Identifiability by block	more than 4 OVs	always identified with the recursive model	more than 4 OVs by block
application fields	sociology, psychologies etc.	marketing, satisfaction etc.	medicine, environment, etc.
Computational time	few minutes	few seconds	few seconds
R packages	lavaan	semPLS	EMsem

Table 1: Comparison between CBSEM, PLS-PM and the EM estimation approach.

This Table (1) allows one to spot advantages of each method and provides indication to pick the approach which is more adapted to one's personal studied case. For example, in the medicine field case, if the user wants to handle treatment (supplementary) covariates in a SEM such as in Model (1) (Barbieri et al, 2017), the EM estimation approach seems to be more appropriated.

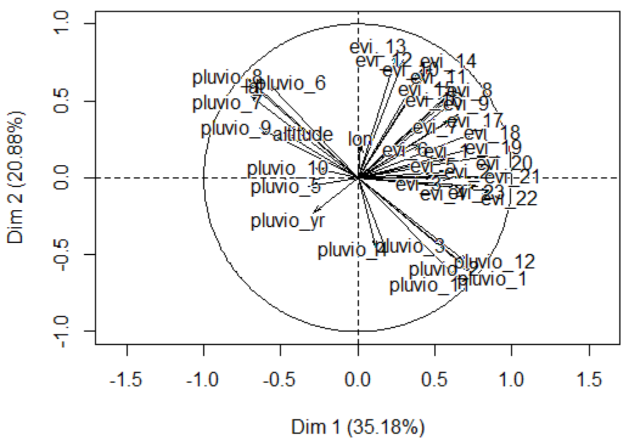

Fig. 7: Correlation-scatterplot yielded by the PCA of the X^1 and X^2 geo-referenced environmental variables (obtained with the FactoMineR R-package).

5 An application to environmental data

5.1 Data presentation

Our model is applied to the data-set *genus*, built from the CoForChange database and provided in the R-package SCGLR developed by Mortier et al (2014). It gives the abundances of 27 common tree genera present in the tropical moist forest of the Congo-Basin, and the measurements of 40 geo-referenced environmental variables, for $n = 1000$ inventory plots (observations). Some of the geo-referenced environmental variables describe 16 physical factors pertaining to topography, geology and rainfall description. The remaining variables characterize vegetation through the enhanced vegetation index (EVI) measured on 16 dates. In this section, we aim at modeling the tree abundances from the other variables, while reducing the dimension of data. The dependent block of variables Y therefore consists of the $q_Y = 27$ tree species counts divided by the plot-surface. A PCA of the geo-referenced environmental variables and the photosynthetic activity variables confirms that EVI measures are clearly separated from the other variables (cf. Figure 7). Indeed, Figure 7 shows two variable-bundles with almost orthogonal central directions. This justifies using our model (cf. Section 5.2) with $p = 2$ explanatory groups, one of them (X^1) gathering $q_1 = 16$ rainfall measures and location variables (longitude, latitude and altitude), and the second one (X^2), the $q_2 = 23$ EVI measures. Besides, in view of the importance of the geological substrate on the spatial distribution of tree species in the Congo Basin, showed by Fayolle et al (2012), we chose to put nominal variable *geology* in a block T. This block therefore contains constant 1 plus all the indicator variables of geology but one, which will therefore be the reference value. Geology having 5 levels, T has 5 columns.

5.2 Model specification with geologic covariates

Here is the model used with the variable-blocks designed in Section 5.1.:

$$\begin{cases} Y &= TD + gb' + \varepsilon^Y \\ X^1 &= \mathbb{1}_n d^{1\prime} + f^1 a^{1\prime} + \varepsilon^1 \\ X^2 &= \mathbb{1}_n d^{2\prime} + f^2 a^{2\prime} + \varepsilon^2 \\ g &= f^1 c^1 + f^2 c^2 + \varepsilon^g \end{cases}$$

where $n = 1000$, $q_Y = 27$, $q_1 = 16$, $q_2 = 23$ and $r_T = 5$. The first row of D is a parameter vector that contains the means of the Y's, and the other rows contain the overall effects of the geological substrates with respect to the reference one. Next section presents the model's parameter-estimates where, in Table 2, each row r of D is noted $D[r,]$.

5.3 Results

With a threshold value $\varepsilon = 10^{-3}$, convergence was reached after 58 iterations. Some parameter-estimates are presented in Tables 2 and 3. For practical reasons, the remaining tables of parameter-estimates are given in the supplementary material.

	Parameter-estimates						
Variables	$D[1,]$	$D[2,]$	$D[3,]$	$D[4,]$	$D[5,]$	b	Correlations with $\widetilde{g}$
gen1	0.76	0.16	0.06	0.68	-0.12	-0.13	-0.14
gen2	0.54	-0.28	-0.03	-0.03	-0.28	0.47	0.58
gen3	0.41	-0.23	-0.02	0.25	-0.37	0.29	0.36
gen4	0.12	0.14	0.03	0.52	0.30	0.09	0.15
gen5	0.31	0.15	0.19	-0.20	0.84	0.09	0.16
gen6	0.55	-0.12	-0.26	0.06	-0.02	0.14	0.18
gen7	0.46	0.06	-0.04	-0.37	0.43	0.14	0.18
gen8	0.55	0.04	-0.09	-0.16	0.04	0.42	0.52
gen9	0.92	-0.54	0.26	-0.66	-0.61	0.07	0.03
gen10	0.68	0.40	0.20	0.37	0.06	-0.32	-0.39
gen11	1.74	-0.50	-0.21	0	-0.67	0.33	0.39
gen12	0.87	0.14	0.73	-0.51	-0.21	0.24	0.26
gen13	1.08	-0.09	-0.37	-0.02	-0.53	0.26	0.29
gen14	0.41	-0.16	-0.10	0.12	-0.36	-0.05	-0.07
gen15	0.51	0.01	-0.11	0.27	-0.18	0.29	0.37
gen16	0.50	-0.19	-0.01	0.55	-0.27	0.1	0.14
gen17	0.79	-0.54	-0.20	-0.52	-0.45	0.39	0.45
gen18	0.16	-0.05	0.20	0.03	-0.03	0.18	0.23
gen19	0.34	0.06	0.41	-0.11	0.38	0.23	0.31
gen20	0.49	0.02	-0.21	0.08	0.14	-0.2	-0.24
gen21	0.79	-0.30	-0.12	0.71	-0.13	0.12	0.19
gen22	0.32	-0.07	-0.07	0.38	-0.11	0.23	0.3
gen23	1.02	-0.28	-0.31	0	-0.07	0.46	0.58
gen24	0.80	-0.23	-0.08	0.22	-0.47	0.57	0.7
gen25	0.60	-0.16	-0.04	0.97	-0.49	0.41	0.53
gen26	0.84	0.22	0.27	-0.70	0.82	0.04	0.07
gen27	0.27	0.41	0.69	-0.24	0.56	0.08	0.11

Table 2: Application to the *genus* data with geologic covariate : estimates of parameters D and b, and correlations of $\widetilde{g}$ with the variables in Y.

Scalar parameter-estimates				
c^1	c^2	σ_1^2	σ_2^2	σ_Y^2
0.35	0.01	0.50	0.53	0.84

Table 3: Application to *genus* data with geologic covariate: scalar parameter-estimates.

It can be seen in Tables 2 and 4 that for certain species, the geologic substrate seems to be of great importance (e.g. for gen1, gen5, gen7, gen9, gen12, gen16, gen21, gen25, gen26, gen27), whereas for others, it only has a small impact on the abundances (e.g. for gen2, gen6, gen8, gen10, gen18, gen20, gen23). Moreover, Table 2 shows that the correlations of $\widetilde{g}$ with Y are high in absolute value only for few variables : *gen2*, *gen23*, *gen24* and *gen25*.

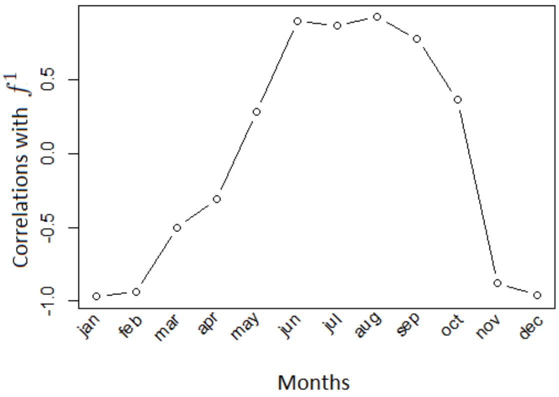

Fig. 8: Correlations of $\widetilde{f^1}$ with the monthly variables of X^1: two rainfall regimes.

Therefore, only these are well accounted for by our model. Although we have carried out the analysis with variables *gen2*, *gen3*, *gen8*, *gen10*, *gen11*, *gen15*, *gen17*, *gen23*, *gen24* and *gen25*, the results are practically the same as when we take all variables. The correlations of $\widetilde{f^1}$ with variables *pluvio_ 1* to *pluvio_ 12* of X^1 show two rainfall regimes (cf. Figure 8). Indeed, *pluvio_ 1* corresponds to january, *pluvio_ 2*, to february, ..., *pluvio_ 12* to december. The Central African Republic has a tropical climate : the dry season ranges from November to April and the rainy season from June to September. Figure 8 shows that $\widetilde{f^1}$ is positively correlated to the rainfalls of the rainy season and negatively to those of the dry one.

5.4 Assessing the model quality through re-sampling

To assess the stability of results and thus, validate the models with covariate, we used a 5-fold re-sampling technique: 5 separate 200 units-samples were randomly extracted from the complete *genus*. For each, we obtained estimated parameters and factors. Then, for each sample, we computed an average Mean Square Error (MSE) and an average correlation of the parameter-estimates obtained on the sample with those obtained on the complete data. Finally, on each sample, we calculated an average MSE and correlation of the factor-estimates obtained on the sample with the corresponding ones obtained on the complete data for the units belonging to the sample.

Figure 9 (resp. 10) shows the average MSE (resp. the correlation) between parameters $\theta_{s\in\{1,\ldots,5\}}$ estimated on 5 data samples (s the index of set of parameters θ associated to the sth data sample) and parameters estimated on the complete data θ. More precisely, for these average MSE (respectively correlations), the median is 3.85×10^{-3} (resp. 0.99), the first quartile is 1.95×10^{-3} (resp. 0.99) and the third quartile is 6.17×10^{-3} (resp. 0.99). These values are close to 0 (resp. 1). Thus, we can be rather confident in the estimates of parameters obtained in the previous section.

Figure 11 and Figure 12 respectively give the box-plot of the factors' average MSE and correlation for each of the 5 samples. More precisely, for these average MSE's (respectively

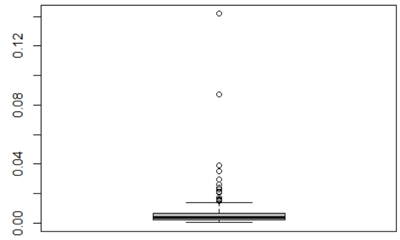

Fig. 9: Box plot of the average MSE's of the parameter-estimates obtained on the 5 *genus* data sub-samples and those obtained on the complete data.

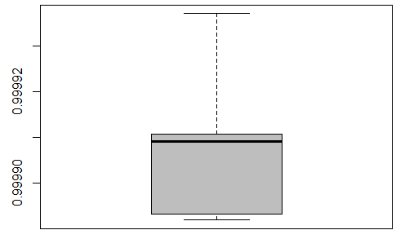

Fig. 10: Box plot of the average correlations of the parameter-estimates obtained on the 5 *genus* data sub-samples and those obtained on the complete data.

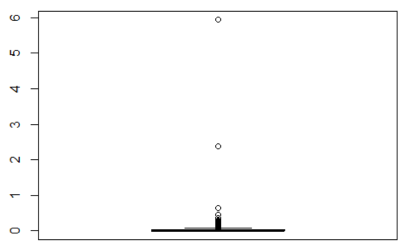

Fig. 11: Box-plot of the average MSE of factor-estimates.

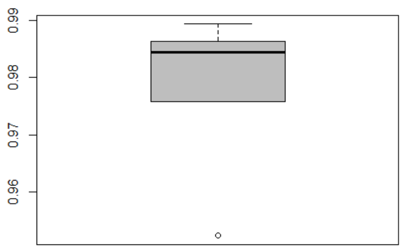

Fig. 12: Box-plot of the average correlation of factor-estimates.

correlations), the median is 1.15×10^{-2} (resp. 0.98), the first quartile is 7.44×10^{-3} (resp. 0.98) and the third quartile is 3.53×10^{-2} (resp. 0.99). These values are close enough to 0 (resp. 1) to allow us to be confident in the estimates obtained on the complete data.

6 Conclusion

The maximum-likelihood estimation method is known to be a stringent method of estimation with nice properties. In the context of estimation methods of a SEM, the CBSEM approach is based on likelihood maximization, contrary to PLS-PM and other component-based methods. However, CBSEM mainly focuses on the variance-covariance structure, the likelihood of which it maximizes under constraints. CBSEM approach does not use the structural equations of the overall SEM: the LVs scores estimation is based on a least squares technique performed exclusively on the mere measurement equations. To estimate both parameters and scores in a row, we proposed to carry out likelihood maximization of the complete model (i.e. both structural equations and measurement equations) through the EM algorithm with a smart initialization step based on PCA which allows avoiding the well-known potential problems of this algorithm. This approach assumes that LVs are factors, which is less constraining than assuming they are components. Therefore, this approach has clear advantage over the more classical ones. And the EM approach and CBSEM can be viewed as complementary methods. Sensitivity analysis allowed to assess its performances. Eventually, the application on environmental data proved satisfactory and demonstrated how to practically use this method. Furthermore the EM estimation approach can handle supplementary covariate such as the geology in the application on environmental data or the treatment in the

medicine case (Barbieri et al, 2017). The implementation of the EM approach is available as an R package on GitHub.
This approach is being numerically compared with lavaan and semPLS R implementations of the CBSEM and PLS-PM methods. Very encouraging preliminary results show that the EM approach provides predictions of latent factors more accurately than PLS-PM and then CBSEM. For the parameters estimation, they seem to be all 3 comparable although for some parameters the comparison seems difficult. Indeed, some parameters are fixed to the unit at CBSEM to allow the identifiability of the model. Other conditions are set for the EM approach. In addition, for some small data sizes, CBSEM does not work while PLS-PM and the EM approach provide good results. Although these results are encouraging, the SEM model for which the EM approach is presented here deserves to be complexifed. It is envisaged to add correlated errors (Schoenberg and Richtand, 1984) and to continue the comparison work between SEM estimation approaches. An idea could be the use of generalized least-squares approach in the M-step such as proposed in (Schoenberg and Richtand, 1984). But, even if experiments with this method has revealed no issues (Schoenberg and Richtand, 1984) precises that this process provides only asymptotically maximum likelihood. Therefore the convergence is not guaranteed. Another perspective of this EM estimation approach, which is a current collaborative work, is about the development of a supplementary step to catch the different explanatory blocks of the OVs in an unsupervised way. Indeed, to construct practically a model and to identify the explanatory blocks of OVs, the use of PCA is proposed in the Section 5. Instead of this manual step, we propose to add an algorithm step which, given an assumed number of explanatory blocks, computes the probability that each explanatory variable belongs to one block.

Appendix A. Calculation of the complete data log-likelihood function $\mathcal{L}$

In the case of the simplified model (2), $p = 2$, $\psi_Y = \sigma_Y^2 I_{q_Y}$, $\psi_1 = \sigma_1^2 I_{q_1}$ and $\psi_2 = \sigma_2^2 I_{q_2}$, and for observation i we have,

$$\begin{aligned}
p(z_i, h_i; \theta) &= p(y_i, x_i^1, x_i^2, g_i, f_i^1, f_i^2; \theta)\\
&= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i, f_i^1, f_i^2; \theta)\\
&= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1, f_i^2; \theta)\\
&= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1; \theta) p(f_i^2; \theta)\\
&= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)\\
&= p(x_i^1, x_i^2 | y_i, g_i, f_i^1, f_i^2; \theta) p(y_i | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)\\
&= p(x_i^1, x_i^2 | f_i^1, f_i^2; \theta) p(y_i | g_i; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)\\
&= p(x_i^1 | x_i^2, f_i^1, f_i^2; \theta) p(x_i^2 | f_i^1, f_i^2; \theta) p(y_i | g_i; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)\\
&= p(x_i^1 | f_i^1; \theta) p(x_i^2 | f_i^2; \theta) p(y_i | g_i; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)
\end{aligned}$$

where $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \psi_Y, \psi_1, \psi_2\}$ is the set of model parameters. Therefore,

$$\mathcal{L}(\theta; z_i, h_i) = \mathcal{L}(\theta; x_i^1 | f_i^1) + \mathcal{L}(\theta; x_i^2 | f_i^2) + \mathcal{L}(\theta; y_i | g_i) + \mathcal{L}(\theta; g_i | f_i^1, f_i^2) + \mathcal{L}(f_i^1) + \mathcal{L}(f_i^2)$$

Because of the model and the normal distribution properties we obtain:
$x_i^m | f_i^m \sim \mathcal{N}(t_i^{m\prime} D^m + f_i^m a^{m\prime}, \psi_{X^m})$
$y_i | g_i \sim \mathcal{N}(t_i{}' D + g_i b', \psi_Y)$
$g_i | f_i^1, f_i^2 \sim \mathcal{N}(f_i^1 c^1 + f_i^2 c^2, 1)$
$f_i^m \sim \mathcal{N}(0, 1)$
Then, we get the complete data log-likelihood function (3), where λ is a constant. Also, the set of model parameters $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \psi_Y, \psi_1, \psi_2\}$ in our case corresponds to $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$ because of the simplification made in the

section 2.3. Indeed, $\psi_Y = \sigma_Y^2 I_{q_Y}$, $\psi_1 = \sigma_1^2 I_{q_1}$ and $\psi_2 = \sigma_2^2 I_{q_2}$.
Therefore, we can also write the complete data log-likelihood function with replacing $ln|\psi_Y|$ (resp. $\forall m \in \{1,2\}$, $ln|\psi_m|$) by $q_Y ln(\sigma_Y^2)$ (resp. $\forall m \in \{1,2\}$, $q_m ln(\sigma_m^2)$).

Appendix B. Distribution of $h_i|z_i$

In the case of the simplified model (2), $p = 2$, $\psi_Y = \sigma_Y^2 I_{q_Y}$, $\psi_1 = \sigma_1^2 I_{q_1}$ and $\psi_2 = \sigma_2^2 I_{q_2}$, and for observation i, the normality of the distribution of $h_i|z_i$ presented in section 3.1.2. derives from the classical result[5] about the conditioning of normally distributed variables. Before using this result, we calculate the joint distribution of $(g_i, f_i^1, f_i^2, y_i, x_i^1, x_i^2)$.

We know that, for observation i,
$y_i \sim \mathcal{N}(D' t_i, b((c^1)^2 + (c^2)^2 + 1)b' + \Psi_Y)$
$x_i^m \sim \mathcal{N}(D^{m\prime} t_i^m, a^m a^{m\prime} + \Psi_m)$
$g_i \sim \mathcal{N}(0, (c^1)^2 + (c^2)^2 + 1)$
$f_i^m \sim \mathcal{N}(0, 1)$
Then, after calculating the required covariances we obtain,

$$(g_i, f_i^1, f_i^2) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}(c^1)^2+(c^2)^2+1 & c^1 & c^2\\ c^1 & 1 & 0\\ c^2 & 0 & 1\end{pmatrix}\right)$$

and,

$$(y_i, x_i^1, x_i^2) \sim \mathcal{N}\left(\begin{pmatrix}D' t_i\\ D^{1\prime} t_i^1\\ D^{2\prime} t_i^2\end{pmatrix}, \begin{pmatrix}((c^1)^2+(c^2)^2+1)bb' + \Psi_Y & c^1 b a^{1\prime} & c^2 b a^{2\prime}\\ c^1 a^1 b' & a^1 a^{1\prime} + \Psi_1 & 0_{(q_1,q_2)}\\ c^2 a^2 b' & 0_{(q_2,q_1)} & a^2 a^{2\prime} + \Psi_2\end{pmatrix}\right)$$

Then, after calculating the required covariances we obtain the joint distribution, $(g_i, f_i^1, f_i^2, y_i, x_i^1, x_i^2) \sim \mathcal{N}(M_i^*, \Sigma^*)$ such that,

$$M_i^* = \begin{pmatrix}0_{(3,1)}\\ D' t_i\\ D^{1\prime} t_i^1\\ D^{2\prime} t_i^2\end{pmatrix} \text{ and } \Sigma^* = \begin{pmatrix}\Sigma_1^* & \Sigma_2^*\\ \Sigma_2^{*\prime} & \Sigma_3^*\end{pmatrix}.$$

Where, $\Sigma_1^* = \begin{pmatrix}(c^1)^2+(c^2)^2+1 & c^1 & c^2\\ c^1 & 1 & 0\\ c^2 & 0 & 1\end{pmatrix}$; $\Sigma_2^* = \begin{pmatrix}((c^1)^2+(c^2)^2+1)b' & c^1 a^{1\prime} & c^2 a^{2\prime}\\ c^1 b' & a^{1\prime} & 0_{(1,q_2)}\\ c^2 b' & 0_{(1,q_1)} & a^{2\prime}\end{pmatrix}$;

$$\Sigma_3^* = \begin{pmatrix}((c^1)^2+(c^2)^2+1)bb' + \Psi_Y & c^1 b a^{1\prime} & c^2 b a^{2\prime}\\ c^1 a^1 b' & a^1 a^{1\prime} + \Psi_1 & 0_{(q_1,q_2)}\\ c^2 a^2 b' & 0_{(q_2,q_1)} & a^2 a^{2\prime} + \Psi_2\end{pmatrix}.$$

Finally, we use result (6) and obtain the distribution, $h_i|z_i \sim \mathcal{N}(M_i, \Sigma)$ where, $M_i = \Sigma_2^* \Sigma_3^{*-1} \mu_i^*$ and $\Sigma = \Sigma_1^* - \Sigma_2^* \Sigma_3^{*-1} \Sigma_2^{*\prime}$, such that $\mu_i^* = \begin{pmatrix}y_i - D' t_i\\ x_i^1 - D^{1\prime} t_i^1\\ x_i^2 - D^{2\prime} t_i^2\end{pmatrix}$.

[5] If two variables X_1 and X_2 are normally distributed such that,
$\begin{pmatrix}X_1\\X_2\end{pmatrix} \sim \mathcal{N}\left(\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$
where, μ_1 $(r \times 1)$, μ_2 $(s \times 1)$, Σ_{11} $(r \times r)$, $\Sigma_{12}(r \times s)$, Σ_{21} $(s \times r)$ and Σ_{22} $(s \times s)$;
then,

$$(X_1|X_2 = x_2) \sim \mathcal{N}(M = \mu_1 + \Sigma_{12}\Sigma_{22}{}^{-1}(x_2 - \mu_2), \phi = \Sigma_{11} - \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}) \qquad (6)$$

Appendix C. Calculation of the first-order derivatives of $\mathcal{L}$ and demonstration of the estimators' formulas of the SEM parameters $\hat{\theta}$

We calculate the first-order derivatives of the complete data log-likelihood function (3), where $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \psi_Y, \psi_1, \psi_2\}$, $\psi_Y = \sigma_Y^2 I_{q_Y}$, $\psi_1 = \sigma_1^2 I_{q_1}$ and $\psi_2 = \sigma_2^2 I_{q_2}$.
There are matrix-parameters (D, D^1, D^2), vector-parameters (b, a^1, a^2) and scalar parameters $(c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2)$. Then, $\mathcal{L}$ is a sum of three types of functions: the logarithm, the square function and a quadratic form function $(w - X\beta)'\Gamma(w - X\beta)$, where Γ is symmetric and w $(q \times 1)$, X $(q \times m)$, β $(m \times 1)$ and Γ $(q \times q)$. The first-order derivatives of the logarithm function and the square function are in our case trivial. The first-order derivative of $(w - X\beta)'\Gamma(w - X\beta)$ with respect to X is less trivial but necessary.

$$
\begin{aligned}
d_X[(w - X\beta)'\Gamma(w - X\beta)] &= (w - X\beta)'\Gamma(-dX\beta) + (-dX\beta)'\Gamma(w - X\beta) \\
&= -2(w - X\beta)'\Gamma(dX\beta) \\
&= tr[-2(w - X\beta)'\Gamma(dX\beta)] \\
&= tr[-2\beta(w - X\beta)'\Gamma dX] \\
&=< -2\beta(w - X\beta)'\Gamma | dX >
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{d}{dX}[(w - X\beta)'\Gamma(w - X\beta)] &= (-2\beta(w - X\beta)'\Gamma)' \\
&= -2(\beta(w - X\beta)'\Gamma)' \\
&= -2\Gamma(w - X\beta)\beta'
\end{aligned}
$$

Likewise, we establish that :

$$
\frac{\partial}{\partial D'}\mathcal{L}(z, h) = \sum_{i=1}^{n} \psi_Y^{-1}(y_i - D't_i - g_i b){t_i}'
$$

Similar reasoning can be applied to D^m and allows to obtain the second row of the following system (7) characterizing all the first-order derivatives of $\mathcal{L}$ with respect to θ:

$$
\begin{cases}
\frac{\partial}{\partial D'}\mathcal{L}(z,h) &= \sum_{i=1}^{n} \psi_Y^{-1}(y_i - D't_i - g_i b){t_i}' \\
\frac{\partial}{\partial D^{m\prime}}\mathcal{L}(z,h) &= \sum_{i=1}^{n} \psi_m^{-1}(x_i^m - D^{m\prime}t_i^m - f_i^m a^m)t_i^{m\prime} \\
\frac{\partial}{\partial b}\mathcal{L}(z,h) &= \sum_{i=1}^{n} g_i\psi_Y^{-1}(y_i - D't_i - g_i b) \\
\frac{\partial}{\partial a^m}\mathcal{L}(z,h) &= \sum_{i=1}^{n} f_i^m\psi_m^{-1}(x_i^m - D^{m\prime}t_i^m - f_i^m a^m) \\
\frac{\partial}{\partial c^m}\mathcal{L}(z,h) &= \sum_{i=1}^{n} f_i^m(g_i - c^2\ f_i^2 - c^1\ f_i^1) \\
\frac{\partial}{\partial \sigma_Y^2}\mathcal{L}(z,h) &= n\ q_Y\ \sigma_Y^{-2} - \sigma_Y^{-4}\sum_{i=1}^{n} ||y_i - D't_i - g_i b||^2 \\
\frac{\partial}{\partial \sigma_m^2}\mathcal{L}(z,h) &= n\ q_m\ \sigma_m^{-2} - \sigma_m^{-4}\sum_{i=1}^{n} ||x_i^m - D^{m\prime}t_i^m - f_i^m a^m||^2.
\end{cases}
\tag{7}
$$

Concerning the third and the fourth row of (7), we use the classical result :

$$
\frac{\partial}{\partial \beta}[(w - X\beta)'\Gamma(w - X\beta)] = -2X'\Gamma(w - X\beta)
$$

Eventually, the fifth, the sixth and the eighth rows of (7) are obtained in a trivial way. To obtain the estimators' formulas of the SEM parameters $\hat{\theta}$, formula (4) (and also (7)) develops into:

$$
\left\{
\begin{array}{rl}
\sum_{i=1}^{n} (y_i - D' t_i - \widetilde{g}_i b) {t_i}' & = 0 \\
\sum_{i=1}^{n} (x_i^m - D^{m\prime} t_i^m - \widetilde{f_i^m} a^m) t_i^{m\prime} & = 0 \\
\sum_{i=1}^{n} \widetilde{g}_i y_i - \widetilde{g}_i D' t_i - \widetilde{\gamma}_i b & = 0 \\
\sum_{i=1}^{n} \widetilde{f_i^m} x_i^m - \widetilde{f_i^m} D^{m\prime} t_i^m - \widetilde{\phi_i^m} a^m & = 0 \\
\sum_{i=1}^{n} \sigma_{12} + \widetilde{f_i^1} \widetilde{g}_i - c^2 \ \sigma_{23} - c^2 \ \widetilde{f_i^1} \widetilde{f_i^2} - \widetilde{\phi_i^1} c^1 & = 0 \\
\sum_{i=1}^{n} \sigma_{31} + \widetilde{f_i^2} \widetilde{g}_i - c^2 \ \widetilde{\phi_i^2} - c^1 \sigma_{32} - c^1 \widetilde{f_i^1} \widetilde{f_i^2} & = 0 \\
n q_Y \sigma_Y^{-2} - \sigma_Y^{-4} \sum_{i=1}^{n} ||y_i - D' t_i||^2 + ||b||^2 \widetilde{\gamma}_i - 2(y_i - D' t_i)' \widetilde{g}_i b & = 0 \\
n q_m \sigma_m^{-2} - \sigma_m^{-4} \sum_{i=1}^{n} ||x_i^m - D^{m\prime} t_i^m||^2 + ||a^m||^2 \widetilde{\phi_i^m} - 2(x_i^m - D^{m\prime} t_i^m)' \widetilde{f_i^m} a^m & = 0
\end{array}
\right.
\tag{8}
$$

System (8) is easy to solve and as an example, the following system characterizing the solution of (4) for the model (2):

$$
\left\{
\begin{array}{l}
\hat{b} = \frac{(\overline{\widetilde{g} y} - \overline{y t'})(\overline{t t'})^{-1} \ \overline{\widetilde{g} t}}{\overline{\widetilde{\gamma}} - \overline{\widetilde{g} t'} (\overline{t t'})^{-1} \overline{\widetilde{g} t}} \\
\widehat{a^m} = \frac{\overline{\widetilde{f^m} x^m} - \overline{x^m t^{m\prime}} (\overline{t^m t^{m\prime}})^{-1} \overline{\widetilde{f^m} t^m}}{\overline{\widetilde{\phi^m}} - \overline{\widetilde{f^m} t^{m\prime}} (\overline{t^m t^{m\prime}})^{-1} \overline{\widetilde{f^m} t^m}} \\
\widehat{c^1} = \frac{(\sigma_{12} + \overline{\widetilde{f^1} \widetilde{g}}) \overline{\widetilde{\phi}^2} - (\sigma_{13} + \overline{\widetilde{f^2} \widetilde{g}})(\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})}{\overline{\widetilde{\phi}^1 \widetilde{\phi}^2} - (\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})^2} \\
\widehat{c^2} = \frac{(\sigma_{13} + \overline{\widetilde{f^2} \widetilde{g}}) \overline{\widetilde{\phi}^1} - (\sigma_{12} + \overline{\widetilde{f^1} \widetilde{g}})(\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})}{\overline{\widetilde{\phi}^1 \widetilde{\phi}^2} - (\sigma_{23} + \overline{\widetilde{f^1} \widetilde{f^2}})^2} \\
\widehat{D'} = (\overline{y t'} - \widehat{b} \ \overline{\widetilde{g} t'})(\overline{t t'})^{-1} \\
\widehat{D^{m\prime}} = (\overline{x^m t^{m\prime}} - \widehat{a^m} \, \overline{\widetilde{f^m} t^{m\prime}})(\overline{t^m t^{m\prime}})^{-1} \\
\widehat{\sigma_Y^2} = \frac{1}{n q_Y} \sum_{i=1}^{n} \{ ||y_i - \widehat{D'} t_i||^2 + ||\hat{b}||^2 \widetilde{\gamma}_i - 2(y_i - \widehat{D'} t_i)' \hat{b} \widetilde{g}_i \} \\
\widehat{\sigma_m^2} = \frac{1}{n q_m} \sum_{i=1}^{n} \{ ||x_i^m - \widehat{D^{m\prime}} t_i^m||^2 + ||\widehat{a^m}||^2 \widetilde{\phi_i^m} - 2(x_i^m - \widehat{D^{m\prime}} t_i^m)' \widehat{a^m} \, \widetilde{f_i^m} \}
\end{array}
\right.
\tag{9}
$$

Appendix D. Table (4) of section 5.

Variables	Differences	Variables	Differences
gen1	0.80	gen15	0.45
gen2	*0.28*	gen16	0.82
gen3	0.62	gen17	0.54
gen4	0.52	gen18	*0.25*
gen5	1.04	gen19	0.52
gen6	*0.32*	gen20	*0.35*
gen7	0.80	gen21	1.01
gen8	*0.20*	gen22	0.49
gen9	0.92	gen23	*0.31*
gen10	*0.40*	gen24	0.69
gen11	0.67	gen25	1.46
gen12	1.24	gen26	1.52
gen13	0.53	gen27	0.93
gen14	0.48		

Table 4: Application to the *genus* data with geologic covariate : Differences between maximal and minimal values of geologic effects $D[1,]$, $D[1,]+D[2,]$, $D[1,]+D[3,]$, $D[1,]+D[4,]$, $D[1,]+D[5,]$ (highlights on the greater differences, italics on the smaller).

Acknowledgements Warmly acknowledgements to V. Brault and E. Devijver for useful discussions, their advises and suggestions for improvement. Then the authors acknowledge the forest inventories were funded by the French Agency for Development (AFD) through the PARPAF project. The authors wish to thank S. Chong (TCA), A. Banos (SCAF), and the ‘"Ministère des Eaux, Forêts, Chasse et Pêche" of the Central African Republic for authorizing access to the inventory data, and the field teams who drew up these inventories. This inventory is part of the ErA Net BiodivERsA CoForChange project, funded by the National Research Agency (ANR) and the Natural Environment Research Council (NERC), involving 16 European, African and international partners and a number of timber companies (see the list on the website, http://www.coforchange.eu).

References

Anderson TW, Rubin H (1956) Statistical inference in factor analysis. In: Proceedings of the third Berkeley symposium on mathematical statistics and probability, vol 5, pp 111–150

Andrade DtF, Helms RW (1984) Maximum likelihood estimates in the multivariate normal with patterned mean and covariance via the em algorithm. Communications in Statistics - Theory and Methods 13(18):2239–2251

Arbuckle J (2011) Ibm spss amos 20 user’s guide: Ibm corporation

Arbuckle JL, Marcoulides GA, Schumacker RE (1996) Full information estimation in the presence of incomplete data. Advanced structural equation modeling: Issues and techniques 243:277

Barbieri A, Tami M, Bry X, Azria D, Gourgou S, Bascoul-Mollevi C, Lavergne C (2017) Em algorithm estimation of a structural equation model for the longitudinal study of the quality of life. Statistics in medicine

Bentler PM (2006) Eqs 6 structural equations program manual

Bentler PM, Tanaka JS (1983) Problems with em algorithms for ml factor analysis. Psychometrika 48(2):247–251

Bollen KA (1989) Structural Equations with Latent Variables. John Wiley & Sons

Browne MW (1974) Generalized least squares estimators in the analysis of covariance structures. South African Statistical Journal 8(1):1–24
Bry X, Verron T (2015) THEME: THEmatic Model Exploration through Multiple Co-Structure maximization. Journal of Chemometrics 29(12):637–647
Bry X, Redont P, Verron T, Cazes P (2012) THEME-SEER: a multidimensional exploratory technique to analyze a structural model using an extended covariance criterion. J Chemometrics 26(5):158–169
Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological) 39(1):1–38
Dempster AP, Rubin DB, Tsutakawa RK (1981) Estimation in Covariance Components Models. Journal of the American Statistical Association 76(374):341–353
Duncan OD (1966) Path analysis: Sociological examples. American journal of Sociology 72(1):1–16
Fayolle A, Engelbrecht B, Freycon V, Mortier F, Swaine M, Réjou-Méchain M, Doucet JL, Fauvet N, Cornu G, Gourlet-Fleury S (2012) Geological Substrates Shape Tree Species and Trait Distributions in African Moist Forests. PLoS ONE 7(8):e42381
Foulley JL (2002) Algorithme EM : Théorie et application au modèle mixte. Journal de la Société française de statistique 143(3-4):57–109
Haavelmo T (1943) The statistical implications of a system of simultaneous equations. Econometrica, Journal of the Econometric Society pp 1–12
Hair JF, Ringle CM, Sarstedt M (2011) Pls-sem: Indeed a silver bullet. Journal of Marketing theory and Practice 19(2):139–152
Hauser RM, Goldberger AS (1971) The treatment of unobservable variables in path analysis. Sociological methodology 3:81–117
Jöreskog KG (1978) Structural analysis of covariance and correlation matrices. Psychometrika 43(4):443–477
Jöreskog K (1973) G.(1973). a general method for estimating a linear structural equation system. Structural equation models in the social sciences pp 85–112
Jöreskog KG (1970) A general method for analysis of covariance structures. Biometrika 57(2):239–251
Jöreskog KG (2000) Latent Variable Scores and their uses. Scientific Software International
Jöreskog KG, Sörbom D (1982) Recent Developments in Structural Equation Modeling. Journal of Marketing Research 19(4):404–416
Keesling J (2016) Maximum Likelihood Approaches to Causal Flow Analysis. phdthesis, University of Chicago
Koopmans T (1945) Statistical estimation of simultaneous economic relations. Journal of the American Statistical Association 40(232):448–466
Lawley DN (1940) Vi.—the estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh 60(1):64–82
Lee SY, Tang NS (2006) Bayesian Analysis of Nonlinear Structural Equation Models with Nonignorable Missing Data. Psychometrika 71(3):541–564
Lohmöller JB (1989) Latent Variable Path Modeling with Partial Least Squares. Springer Science & Business Media
McDonald RP (1996) Path Analysis with Composite Variables. Multivariate Behavioral Research 31(2):239–270
Mortier F, Trottier C, Cornu G, Bry X (2014) SCGLR-An R Package for Supervised Component Generalized Linear Regression. Journal of Statistical Software
Muthén L, Muthén B (2010) Version 6 mplus user's guide. Los Angeles, CA: Muthén & Muthén
Muthén B, Kaplan D, Hollis M (1987) On structural equation modeling with data that are not missing completely at random. Psychometrika 52(3):431–462
Muthén L, Muthén B (1998) MPlus User's Guide
Rabe-Hesketh S, Skrondal A, Pickles A (2004) Generalized multilevel structural equation modeling. Psychometrika 69(2):167–190
Rubin DB, Thayer DT (1982) Em algorithms for ml factor analysis. Psychometrika 47(1):69–76

Rubin DB, Thayer DT (1983) More on em for ml factor analysis. Psychometrika 48(2):253–257
Saltelli A, Chan K, Scott EM, et al (2000) Sensitivity analysis, vol 1. Wiley New York
Schoenberg R, Richtand C (1984) Application of the em method: A study of maximum likelihood estimation of multiple indicator and factor analysis models. Sociological Methods & Research 13(1):127–150
Spearman C (1904) " general intelligence," objectively determined and measured. The American Journal of Psychology 15(2):201–292
Stan V, Saporta G (2006) Une comparaison expérimentale entre les approches PLS et LISREL. In: 38 èmes Journées de Statistique, Clamart, France, X, France
Tami M (2016) Approche EM pour modèles multi-blocs à facteurs à une équation structurelle. phdthesis, University of Montpellier
Tang ML, Lee SY (1998) Analysis of structural equation models with censored or truncated data via EM algorithm. Computational statistics & data analysis 27(1):33–46
Wiley D, Seminar Press I, of America US (1973) The identification problem for structural equation models with unmeasured variables, structural equation models in the social science. 69-83
Wold H (1966) Estimation of Principal Components and Related Models by Iterative Least squares. In: Multivariate Analysis., Academic Press, pp 391–420
Wold H (1973) Nonlinear iterative partial least squares (nipals) modelling : Some current developments. In: Multivariate Analysis, vol III, Academic Press, pp 383–407
Wold H (1982) Soft modelling: the basic design and some extensions. Systems under indirect observation, Part II pp 36–37
Wold H (1985) Partial Least Squares. In: Encyclopedia of Statistical Sciences, vol 6, John Wiley & Sons, New York, pp 581–591
Wright S (1921) Correlation and causation. Journal of agricultural research 20(7):557–585
Zellner A (1970) Estimation of regression relationships containing unobservable independent variables. International Economic Review pp 441–454