Statistical Methods & Applications
https://doi.org/10.1007/s10260-025-00783-3

ORIGINAL PAPER

Robustification of structural equation modelling via global sensitivity analysis

Alessio Lachi[1] · Josep Llach[3] · Jordi Perramon[3] · Michela Baccini[2] · Andrea Saltelli[3]

Accepted: 14 February 2025

Abstract
We propose a method for enhancing the robustness of Structural Equation Modelling (SEM), a multivariate statistical analysis technique employed for analyzing causal relationships among different aspects of given phenomena. This enhancement is achieved through the integration of Global Sensitivity Analysis, which assesses how uncertainties in model output can be attributed to various sources of input uncertainty. The robustification process involves several key steps, including bootstrapping evidence, error propagation, and uncertainty quantification. This method extends the approach named in the literature "modeling of the modelling process". To illustrate this approach, we apply it to two previously published test cases where SEM is used. The first one is related to the impact of artificial intelligence adoption on employee engagement and the second one investigates the effects of service quality and environmental practices on the competitiveness and financial performance of hotels. By quantifying the uncertainty inherent in the inference of our test cases, this procedure increases the robustness of the results derived from the test cases, thus generating a more defensible inference.

Keywords Structural equations modelling · Global sensitivity analysis · Uncertainty modelling · Uncertainty quantification · Sobol indexes · Robustification · Bootstrap

1 Background

The purpose of this paper is to propose a robustification of structural equation modelling (SEM), a key method used in the social sciences (Wolfle (1982); Wolfle and Ethington (1985); Lomax (1983)). The robustification is achieved by modelling the assumption used in the data analysis, by allowing these assumptions to vary so that an uncertainty and sensitivity analysis (SA) can be performed on these variations. The approach, proposed by Piano et al. (2022, 2023), is known as "modelling of the modelling process". It is based on subjecting the various stages of a model-building

Extended author information available on the last page of the article

Published online: 11 March 2025

process to a coordinated and simultaneous variation in the modelling assumptions. This generates an error propagation analysis (uncertainty analysis, UA), followed by a global sensitivity analysis (GSA, Saltelli et al. (2008)). The key steps of the analysis are the data selection and the attribution of items to latent variables. The mechanic of the analysis entails the use of triggers that are selected at runtime, leading to different paths being taken by the analysis in different simulations. The input data are also bootstrapped, as proposed by Tibshirani and Efron (1993), in each simulation, following the philosophy of "bootstrapping of the modelling process" (Chatfield (1995)). The relevance of this analysis is revealed by experiments such as that discussed by Breznau, N., et al. (2022), where many teams (73 in fact) were given the same data and returned wildly different conclusion, non only on the value of a statistical effect, but also on its sign. What made the experiment of Breznau, N., et al. (2022) relevant for modellers and statisticians is that the teams differed in their work in apparently inconsequential aspects of the treatment of the data, and yet the combined effects of these modelling choices resulted dramatic. An author, commenting on this experiment, talks of "A universe of uncertainty hiding in plain sight" (Engzell (2023)). Statistician Andrew Gelman (Gelman and Loken 2013; Steegen et al. 2016) compares the statisticians running their analysis to subjects walking through a "garden of forking path" - Gelman makes reference here to a short fiction of writer Jorge Luis Borges (Borges 1998).

GSA examines model output sensitivity when varying all uncertain inputs across their entire range simultaneously (Saltelli (2002)), helping modellers in exploring the impact of assumptions, identifying influential inputs, simplifying problems, and supporting model-based decision-making (Razavi et al. (2021)). While GSA is customarily used by sampling uncertain factors from distribution derived from the calibration or by experts, GSA can also sample assumptions made in the course of the analysis, thus simulating the present of different teams tackling the same problem. In other words, in the context of the present work, our GSA corresponds to a stroll in the garden of forking paths.

The most widespread GSA algorithms are variance-based methods, which decompose the output variance of a k-dimensional model into contributions from individual inputs, pairs of inputs, and so on with terms of increasing dimensionality till a k-th term, see (Saltelli et al. 2008) for an introduction. The analysis leads to the estimation of so-called Sobol' indices (Sobol' 1993).

Other GSA procedures are also available. When the output has extreme values, moment-independent methods may be preferred over variance-based methods because they do not assume normality or finite output variance (Borgonovo (2007)). Another approach is the PAWN index, which differs from other moment-independent approaches in its reliance on cumulative distribution functions to characterize the output variance (Pianosi and Wagener (2015)). The use of VARS and Shapley effects has also become more widespread in recent years (Razavi and Gupta (2016)). VARS are based on the use of variogram and covariogram functions to characterize the variability in the output at different scales. Instead, Shapley effects rely on Shapley values, which represent the average marginal contribution of a given feature/factor across all possible feature/factor combinations (Owen (2014)). Also, the concept of Kernel-based dependence has spread in recent years. This concept is based on the definition of the maximum mean discrepancy between the unconditional and conditional output distributions (Barr and Rabitz (2022, 2023)). Despite this abundance of methods, the uptake of GSA in mathematical modelling in still in progress. "One variable at a time" (OAT) procedures, whereby individual factors are perturbed while keeping the other constants, are still very popular, due to their ease of interpretation (Pianosi et al. (2016); Razavi et al. (2021)). The problem with OAT methods is that they do not explore efficiently the space of the input and do not capture interaction effects (Saltelli et al. (2010)).

In the literature, Lee and Wang (1996) proposes for SEM a SA based on first-order derivatives, which, however, only captures the local influence of small perturbations. Instead, Pek and MacCallum (2011) analyze the uncertainty of model estimates under different subsets of the data, while Wang and Lee (1996) explore the sensitivity analysis (SA) of SEM with functional equality constraints when minor perturbations are introduced. More generally, in the context of graphical models, De Bock et al. (2014) examine the sensitivity of the maximum a posteriori configuration of these models to parameter perturbations. Furthermore, Ballester-Ripoll and Leonelli (2021) employ Sobol's method of GSA to quantify the influence of evidence from a set of nodes on a quantity of interest.

Instead, our contribution provides modellers engaged in SEM and similar methods with a direct GSA tool, which allows them to explore the robustness of their inference to change in the assumptions. The proposed procedure is able to detect more influential parameters/assumptions as well as their interactions with one another. With this approach, if for example these uncertainty explorations reveal that a parameter is not influential in defining a model, researchers can create simpler models that achieve the same results.

The present investigation concerns two existing test cases. The first one is related to artificial intelligence (AI) adoption and its relationship with employee engagement explored in Theben et al. (2023). Specifically, this test case considered here involves the mediating role of *training* provision in the relationship between *AI adoption* and three dimensions of employee engagement, *vigour*, *dedication*, and *absorption*, considering job *complexity* as a critical factor. The second case concerns the effects of service quality and environmental practices on the competitiveness and financial performance of hotels. More precisely, this test case focuses on studying the combined effects of quality management practices (*QMP*) and environmental management practices (*EMP*) and their relationship with factors such as competitiveness (*CO*) and financial performance (*FP*), specifically in the tourism sector of hotel companies.

Our replication-cum-uncertainties of the analysis in Theben et al. (2023) and Perramon et al. (2022) simultaneously tests whether the original inference is reproduced while illustrating the applicability of the method to SEM in general. The present work feeds into the current debate over the fragility of model-based inference to modelling assumptions discussed above in relation to Gelman and Loken (2013), Breznau, N., et al. (2022), and Cantone and Tomaselli (2024); see also Saltelli and Di Fiore (2023) for a broader discussion of the topic in the context of modelling.

Our analysis broadly confirms the narrative of Theben et al. (2023) and Perramon et al. (2022) with some relevant and interesting differences in specific findings. It should also be noted that the GSA also offers a richer insight into the relative influence of the items on the construction of the latent variables.

After a brief introduction to the data sources used in the analysis (Sect. 2.1), we describe the SEM approach (Sect. 2.2.1) and illustrate the uncertainty propagation of the error (Sect. 2.2.2). We then report our findings (Sect. 3), discuss them highlighting the merits and limitations of the study and finally draw our conclusions (Sect. 4).

2 Data and methods

2.1 Data

Test case 1

The data collection used in Theben et al. (2023) was conducted in 2022. The questionnaire was structured into two sections. The first section included queries related to the profile of the respondent. The second section contained questions related to the respondents' perception of the *training* offered by their companies and to what extent the respondents felt engaged with their jobs. For all the variables, the authors in Theben et al. (2023) used a seven-level Likert scale Likert (1932). The initial sample size included 302 individuals. A key question concerned the respondent's opinion on whether AI adoption was strategic for the organization. Retaining only participants who considered AI adoption strategic for their companies, a sample of 211 employees was obtained. The items used in Theben et al. (2023) are listed in the Table 1.

An exploratory factor analysis (EFA), as proposed by Fontaine (2005), was performed in Theben et al. (2023) to check the psychometric validity of the questionnaire and of the items related to the dimensions. Exploratory factor analysis is a statistical technique used to reduce data to a smaller set of summary variables and explore the underlying theoretical structure (Norris and Lecavalier (2010)) of the phenomenon being considered. EFA tries to uncover the underlying structure of a relatively large set of variables (Fabrigar et al. (1999)). Five independent EFAs, one for each class of question reported in Table 1, were done to identify the underlying relationships between measured variables (Finch and West (1997)).

Theben et al. (2023) identify the following set of latent variables, constructed under a battery of related items as follows:

- A latent variable that measures the *AI adoption* in a company and is defined by four items $\{14_1, 14_2, 14_3, 14_5\}$;
- A latent variable that measures the *vigour* of a worker and is defined on three items $\{17_1, 17_2, 17_3\}$;
- A latent variable that measures the *dedication* of a worker and is defined on four items $\{17_7, 17_8, 17_9, 17_{10}\}$;
- A latent variable that measures the *absorption* of a worker and is defined on four items $\{17_{12}, 17_{14}, 17_{15}, 17_{16}\}$;

Table 1 Questionnaire used in the analysis of Theben et al. (2023)

Question	Code	Item
Opinions for AI adoption	11_1	AI adoption is strategic for my organization
AI adoption aspect	14_1	My work can be completed by a zero-hour contractor such as a computer program
	14_2	My work may be performed by someone with variable compensation such as compensation based on the performance or results of the work
	14_3	Payment for my work can be made on the basis of previously agreed results
	14_4	My work can be done by someone who is paid by the hour
	14_5	My work can be completed for a lower rate of pay
Vigour aspect	17_1	I feel full of energy at work
	17_2	I feel strong and energetic in my work
	17_3	When I get up in the morning I feel like going to work
	17_4	I can continue to work for long periods
	17_5	I am very persistent in my work
	17_6	Even when things don't go well, I keep working
Dedication aspect	17_7	My work is full of meaning and purpose
	17_8	I am enthusiastic about my work
	17_9	My work inspires me
	17_{10}	I am proud of the work I do
	17_{11}	My work is challenging
Absorption aspect	17_{12}	Time flies when I'm working
	17_{13}	When I am working, I forget everything that is going on around me
	17_{14}	I am happy when I am absorbed in my work
	17_{15}	I am immersed in my work
	17_{16}	I let myself be carried away by my work
	17_{17}	I find it difficult to disconnect from my work

Table 1 (continued)

Question	Code	Item
Training aspect	20_1	The training received is of high quality
	20_2	Training is constantly reviewed and updated to meet the requirements of the changing work environment
	20_3	Experienced employees receive regular training and training updates
	20_4	Experienced employees receive training when new initiatives are launched
	20_5	Employees receive sufficient training opportunities
	20_6	Employees receive training on a systematic basis
	20_7	Newly hired employees receive adequate training
	20_8	The training provided in my organization exceeds the skill requirements to perform my job duties
Complexity aspect	15_1	I am assigned extraordinary and particularly difficult tasks

- A latent variable that measures the *training* of a worker and is defined on four items $\{20_2, 20_3, 20_5, 20_6\}$;
- A latent variable that measures the work *complexity* of a company and is defined on only one item $\{16_1\}$.

More information on the composition of the latent variables and on the items composing them is reported in Theben et al. (2023).

Test case 2

The data used in Perramon et al. (2022) were collected in 2017 and are part of a wider sample used in a previous work by Bagur-Femenías et al. (2019). The survey was randomly distributed to over 1,000 hotels via email and phone calls, aiming to receive a response from one manager per hotel. This resulted in an approximate valid response rate of 14%. The questionnaire was structured into two sections. The first section included queries related to the profile of the respondent. The second section contained questions related to the four constructs explored in the study: *QMP*, *EMP*, *CO*, and *FP*. For all the variables, the authors in Bagur-Femenías et al. (2019) used a five-level Likert scale (Likert (1932)). The initial sample size included 148 individuals. The items used in Perramon et al. (2022) are listed in the Table 2.

As done for the test case 1, an EFA was performed in Perramon et al. (2022) to check the psychometric validity of the questionnaire and of the items related to the dimensions. Four independent EFAs, one for each class of question reported in Table 2, was done to identify the underlying relationships between measured variables (Finch and West (1997)).

Perramon et al. (2022) identify the following set of latent variables, constructed under a battery of related items as follows:

- A latent variable that measures the *QMP* in an hotel and is defined by four items $\{QMP_1, QMP_2, QMP_3, QMP_4\}$;
- A latent variable that measures the *EMP* of an hotel and is defined on three items $\{EMP_1, EMP_2, EMP_3, EMP_4\}$;
- A latent variable that measures the *CO* of an hotel and is defined on four items $\{CO_1, CO_2, CO_3, CO_4\}$;
- A latent variable that measures the *FP* of an hotel and is defined on four items $\{FP_1, FP_2, FP_3\}$.

More information on the composition of the latent variables and on the items composing them is reported in Perramon et al. (2022).

Table 2 Questionnaire used in the analysis of Perramon et al. (2022)

Question	Code	Item
Quality management practices (*QMP*)	QMP_1	All the staff are involved in the creation of the product /service
	QMP_2	Improvements have been identified in the service provision process
	QMP_3	Goal achievement control is conducted, and variations are amended
	QMP_4	There is a quality culture with a focus on continuous improvement
Environmental management practices (*EMP*)	EMP_1	The company quantifies environmental savings
	EMP_2	The company uses ecological factors in marketing campaigns
	EMP_3	The company has a long-term environmental strategic focus
	EMP_4	The company uses a green criterion in its purchasing policy
Competitiveness (*CO*)	CO_1	Improved market image of the facilities
	CO_2	Client satisfaction level is greater than among the competition
	CO_3	Employee satisfaction level is greater than among the competition
	CO_4	Improved capacity to stay in the market in times of crisis
Financial performance (*FP*)	FP_1	Sales have increased in the last two years
	FP_2	Profits have increased in the last two years
	FP_3	Occupation has increased in the last two years

2.2 Methods

2.2.1 Structural equations modelling

SEM is a well-established statistical method for investigating complex relationships between latent constructs. SEM's ability to simultaneously estimate direct, indirect (e.g., mediating), and moderating effects of multiple constructs while accounting for measurement error has enabled researchers to examine relationships that would otherwise be difficult to disentangle (Ringle et al. (2020)). SEM has two main approaches, covariance-based SEM (CB-SEM) and partial least squares-based SEM (PLS-SEM). Schuberth et al. (2023) provide a revision on the structural parameters of SEM under the two approaches. Theben et al. (2023) applied the CB-SEM approach using the software EQS 6.4 Bentler and Wu (2015) based on the maximum likelihood estimation method.

According to Kaplan (2008), SEM can be defined as a class of methodologies aimed at testing hypotheses about the means, variances, and covariances of observed data in terms of a smaller number of structural parameters defined by a hypothesized underlying model. Thus, SEM, often called linear structural relations models, tries to explain relations between latent variables. The main advantage of SEM is its flexibility to deal not only with a single simple or multiple linear regression but also with several equations simultaneously as discussed in Nachtigall et al. (2003).

Test case 1

Let us denote with:

- $M1 = (M^1_{AI\ adoption}, M^1_{vigour}, M^1_{dedication}, M^1_{absorption}, M^1_{training}, M^1_{complexity})$ the matrix of data used in the analysis of Theben et al. (2023);
- $X^1 = (X^1_1, X^1_2, X^1_3) = (M^1_{complexity}, M^1_{training}, M^1_{AI\ adoption})$ represent the matrix of independent variable;
- $Y^1 = (Y^1_1, Y^1_2, Y^1_3, Y^1_4) = (M^1_{training}, M^1_{vigour}, M^1_{dedication}, M^1_{absorption})$ represent the matrix of dependent variable.

To note that for $M^1_{training}$ we used a double notation: X^1_2 if we considered it as independent variable and Y^1_1 if we considered it as dependent variable.

The model defined in Theben et al. (2023), graphically represented in Fig. 1, is based on the following system of equations:

$$\begin{cases} Y^1_1 = \beta^1_3 X^1_3 \\ Y^1_2 = \gamma^1_1 X^1_1 + \gamma^1_2 X^1_2 + \gamma^1_3 X^1_3 \\ Y^1_3 = \omega^1_1 X^1_1 + \omega^1_2 X^1_2 + \omega^1_3 X^1_3 \\ Y^1_4 = \alpha^1_1 X^1_1 + \alpha^1_2 X^1_2 + \alpha^1_3 X^1_3. \end{cases} \tag{1}$$

Test case 2

Let us denote with:

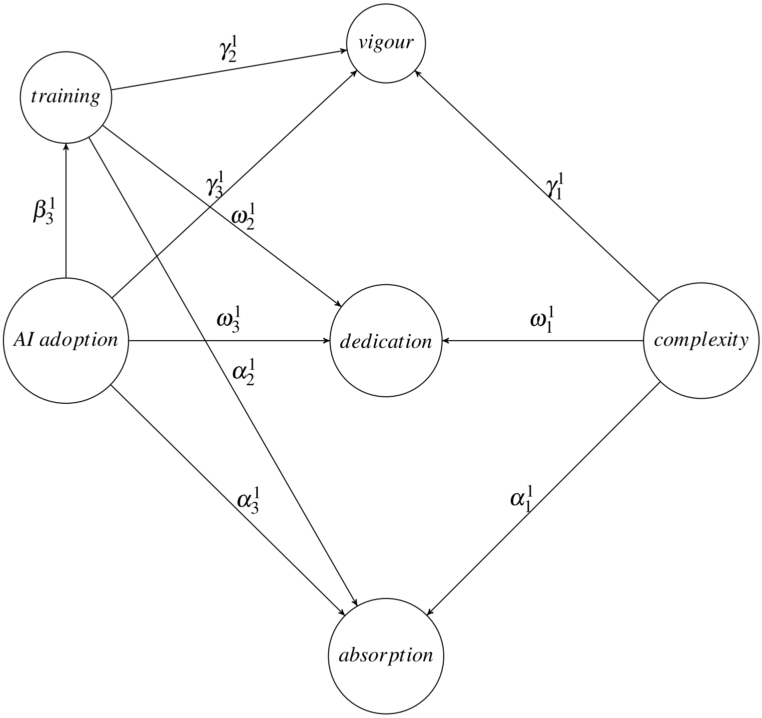

Fig. 1 Graphical representation of the SEM of Theben et al. (2023)

- $M^2 = (M^2_{QMP}, M^2_{EMP}, M^2_{CO}, M^2_{FP})$ the matrix of data used in the analysis of Perramon et al. (2022);
- $X^2 = (X^2_1, X^2_2, X^2_3) = (M^2_{QMP}, M^2_{EMP}, M^2_{CO})$ represent the matrix of independent variable;
- $Y^2 = (Y^2_1, Y^2_2, Y^2_3) = (M^2_{EMP}, M^2_{CO}, M^2_{FP})$ represent the matrix of dependent variable.

To note that:

- For M^2_{QMP} we used a double notation: X^2_2 if we considered it as independent variable and Y^2_1 if we considered it as dependent variable;
- For M^2_{CO} we used a double notation: X^2_3 if we considered it as independent variable and Y^2_2 if we considered it as dependent variable.

The model defined in Perramon et al. (2022), graphically represented in Fig. 2, is based on the following system of equations:

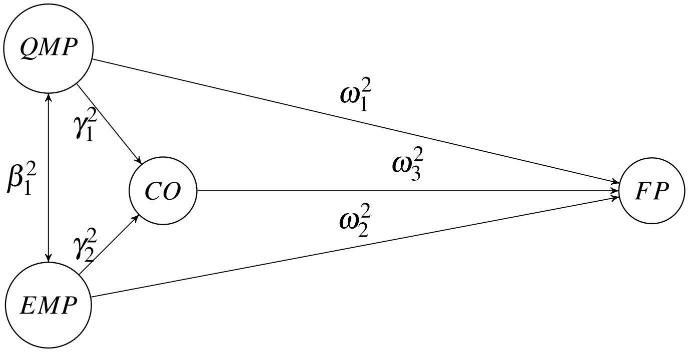

Fig. 2 Graphical representation of the SEM of Perramon et al. (2022)

$$\begin{cases} Y_1^2 = \beta_1^2 X_1^2 \\ Y_2^2 = \gamma_1^2 X_1^2 + \gamma_2^2 X_2^2 \\ Y_3^2 = \omega_1^2 X_1^2 + \omega_2^2 X_2^2 + \omega_3^2 X_3^2. \end{cases} \tag{2}$$

2.2.2 Uncertainty assessment

In statistical modelling, uncertainty assessment is the process of quantifying the uncertainty associated with model assumptions, parameters, and predictions. Uncertainty arises from various sources, such as measurement errors, sampling variability, model misspecification, and unobserved variables.

Usually, a confidence interval for the inference is estimated, that provides range of plausible values for the parameter or prediction, along with a level of confidence in the interval's coverage, based on some parametric assumption distributions. If estimating confidence intervals is difficult, we can base our inference on the bootstrap procedure. The basic idea behind the bootstrap procedure proposed by Tibshirani and Efron (1993) is to create many random samples, with replacements, from the original sample of data. Each of these resamples has the same size as the original sample, and they are drawn independently from the original data. A statistic of interest, such as the mean or the standard deviation, is then calculated for each of these resamples. By repeating this resampling and calculation process many times, a distribution of the statistic can be estimated. This distribution represents the variability of the statistic and can be used to construct confidence intervals or to perform hypothesis tests.

It is appropriate in modelling studies to consider multiple candidate models that differ in their assumptions or specifications. A particularly radical version of this approach is the "modelling of the modelling process", a specific type of GSA (Saltelli et al. (2008)) described in Piano et al. (2022) and Saltelli and Puy (2023). This exploration can be performed in a Monte Carlo (MC) framework, as discussed in Kroese et al. (2014), by defining random "triggers" that determine the model to be followed in each simulation. In this way, we can combine the predictions from these models while accounting simultaneously for the uncertainty associated with each model's parameters.

UA works in tandem with SA, which involves examining the relative importance of model inputs or assumptions in determining the uncertainty in the model outputs. This can help identify which model assumptions are most critical for the results and quantify the impact of potential sources of uncertainty, as discussed by Leite et al. (2022). We investigated the influence of each trigger on the model output, which in this case is represented by the estimates of the coefficients of the SEM, through the total sensitivity index S_{T_i} provided by GSA, a variance-based approach explained in Saltelli et al. (2008).

Given K_Z mutually independent inputs $(Z_1, Z_2, \dots, Z_i, \dots, Z_{K_Z})$ and a model which, given the inputs, returns K_W outputs $(W_1, W_2, \dots, W_j, \dots W_{K_W})$, this approach quantifies the relative importance of each input to the model's outcomes by propagating uncertainty from the inputs to the outputs and computing variance indices. Then we calculated, for each input Z_i, where $i \in \{1, 2, \dots, K_Z\}$, the so-called total variance index, which $S^{tot}_{i,j}$ measures the overall effect of the i-th input on the output W_j, where $j \in \{1, 2, \dots, K_W\}$, including all the interactions of Z_i with the other inputs. This index corresponds to the expected variance of W_j that would be left on average when all the parameters but Z_i, $Z_{\sim i}$, are fixed:

$$S^{tot}_{i,j} = \frac{E_{Z \sim i}(Var_{Z_i}(W_j | Z_{\sim i}))}{Var(W_j)}. \tag{3}$$

The "on average" in the preceding sentence refers to the fact that $S^{tot}_{i,j}$ is averaged over all possible combinations of value fixed for the set of $Z_{\sim i}$ as discussed by Saltelli et al. (2008). A total variance index close to zero indicates that the parameter Z_i does not influence W_j. Instead, a large total variance index indicates that the parameter does have an impact on them. The estimation procedure for $S^{tot}_{i,j}$ is given in Saltelli et al. (2010).

For generating total indices, we define two design matrix A and B of dimension $n \times p$. For the MC simulation, we used quasi-random numbers defined in [0, 1], which allow for a more efficient exploration of the sample space compared to the uniform distribution, as suggested by Kucherenko et al. (2015) and Sobol (1994). Note that the choice of a distribution defined in [0, 1] ensures that each model is equally probable to be sampled. In A and B, p represents the number of inputs of the model coming to play into GSA and n represents the different combinations of model input. In addition to the definitions of matrices A and B, the calculation of sensitivity indexes in GSA requires the definition of the matrix A_i^B with $i \in \{1, \dots, p\}$. This matrix is identical to A, except for column i, which is replaced by column i of matrix B (Saltelli et al. (2008)). In GSA studies the choice of n for the matrices A and B depends on the MC error observed in the simulations. In the test cases presented here, we set the dimension of these matrices to 1,000 because

Table 3 Model estimates of Equation (1) provided by asymptotically inference and bootstrap with their 95% percentage confidence intervals (* stands for a p-value < 0.05)

Coefficient	Asymptotically	Bootstrap
β_3^1	0.090 (−0.090–0.270)	0.091 (−0.049–0.244)
γ_1^1	0.807* (0.706–0.909)	0.808* (0.696–0.891)
γ_2^1	0.408* (0.255–0.562)	0.407* (0.244–0.515)
γ_3^1	−0.131 (−0.294–0.033)	−0.129 (−0.290–0.032)
ω_1^1	0.758* (0.677–0.839)	0.761* (0.682–0.839)
ω_2^1	0.436* (0.309–0.563)	0.431* (0.314–0.518)
ω_3^1	−0.185* (−0.321––0.050)	0.181* (−0.328––0.006)
α_1^1	0.794* (0.684–0.904)	0.800* (0.684–0.882)
α_2^1	0.325* (0.172–0.479)	0.309* (0.182––0.437)
α_3^1	−0.013 (−0.162–0.136)	−0.014 (−0.159–0.153)

we observed that this size is sufficient for adequately exploring the model sample space and provide acceptable MC error in the sensitivity indices.

The focus of this paper is to test the robustness of the SEM-based latent variables using GSA and the "Modelling of the modelling process" just described. The approach includes adding or removing items to the composition of various latent variables and studying how these different compositions of latent variables impact the estimation of the SEM coefficients.

3 Results

All the analyses performed in this paper, were conducted using R 4.4.2, specifically employing the lavaan package for SEM and the sensitivity package for GSA.

Test case 1

As a first step of our analysis, we compared, in the following Table 3 and Fig. 3, model estimates of Equation (1) provided by asymptotically inference with the confidence intervals derived from a bootstrap procedure repeated for one-thousand of times (Tibshirani and Efron (1993)).

Table 3 shows that central points of model estimates of Equation (1), with their 95% confidence intervals provided by asymptotically inference are coherent with the one obtained from bootstrap. In general, the dimensions *vigour*, *dedication*, and *absorption* are better explained by the dimension *complexity*, and the relationship with *AI adoption* is non-significant. The model shows a significant positive effect of training quality and quantity on engagement, in all three of its dimensions. While higher levels of *AI adoption* in a company could be associated with lower training quality and quantity, this hypothesis is rejected as the relationship is non-significant. Instead, higher levels of work complexity were associated with higher vigour and absorption but did not show a significant relationship to the dedication dimension of engagement.

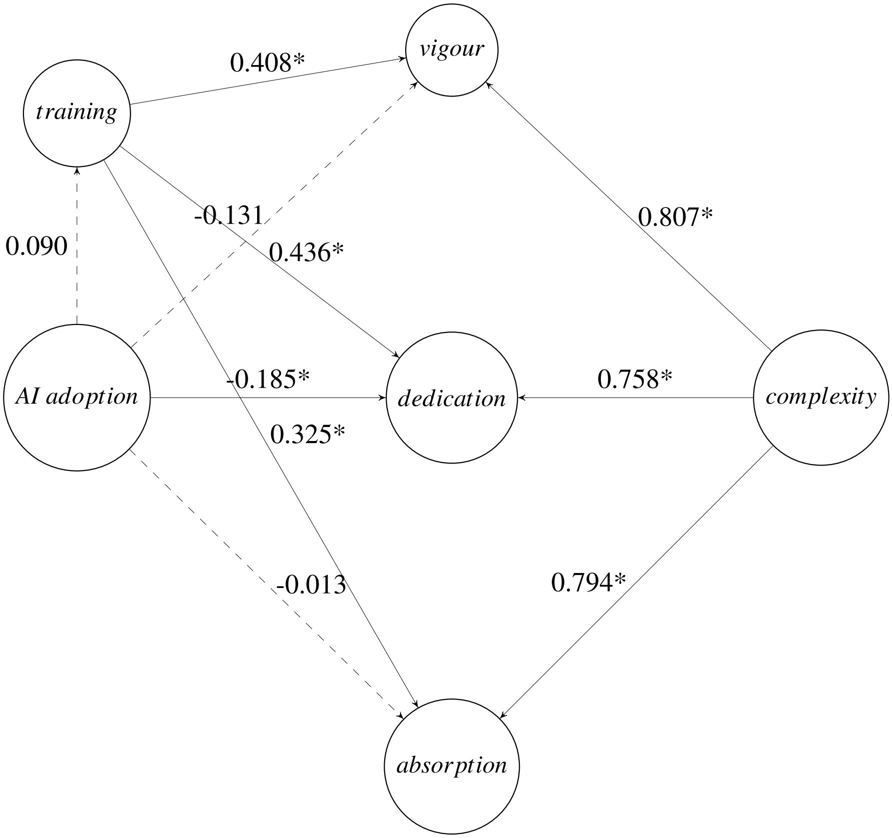

Fig. 3 Graphical representation of the SEM of Theben et al. (2023). Above the arrows, we report the regression coefficient (* stands for a p-value < 0.05). Dashed arrow represent non-significative coefficients

To note that the analysis performed in Theben et al. (2023) considered only individuals who considered AI adoption to be strategic for their companies. As a second step of our analysis, we performed a simple counterfactual analysis to verify the effect of including or not including individuals who consider AI adoption to be not strategic. The following table shows the SEM coefficient estimates produced by considering only individuals who consider AI adoption strategic and only individuals who consider AI adoption to be non-strategic for their companies:

The results reported in Table 4 show that the coefficient estimates of the two groups of individuals are very similar. The regression $Y_1^1 = \beta_3^1 X_3^1$ remain stable. In the regression $Y_2^1 = \gamma_1^1 X_1^1 + \gamma_2^1 X_2^1 + \gamma_3^1 X_3^1$, we note that the inclusion of individuals who consider AI adoption non-strategic make the relation with γ_2^1 non-significative. The regression $Y_3^1 = \omega_1^1 X_1^1 + \omega_2^1 X_2^1 + \omega_3^1 X_3^1$ remain stable. In the regression $Y_4^1 = \alpha_1^1 X_1^1 + \alpha_2^1 X_2^1 + \alpha_3^1 X_3^1$, we note that the inclusion of individuals who consider AI adoption non-strategic make the relation with α_2^1 non-significative. In general,

Table 4 Model estimates of Equation (1) provided by asymptotically inference with their 95% percentage confidence intervals. The model estimates compare individuals who considered AI adoption to be and not to be strategic for their companies (* stands for a p-value < 0.05)

Coefficients	AI adoption strategic	AI adoption not strategic
β_3^1	0.090 (−0.090–0.270)	0.100 (−0.147–0.346)
γ_1^1	0.807* (0.706–0.909)	0.870* (0.762–0.973)
γ_2^1	0.408* (0.255–0.562)	0.180 (−0.037–0.389)
γ_3^1	−0.131 (−0.294–0.033)	−0.110 (−0.350–0.132)
ω_1^1	0.758* (0.677–0.839)	0.830* (0.717–0.937)
ω_2^1	0.436* (0.309–0.563)	0.200* (0.009–0.392)
ω_3^1	−0.185* (−0.321––0.050)	−0.320* (−0.543––0.103)
α_1^1	0.794* (0.684–0.904)	0.860* (0.757–0.965)
α_2^1	0.325* (0.172–0.479)	0.070 (−0.131–0.277)
α_3^1	−0.013 (−0.162–0.136)	−0.210 (−0.431–0.006)

the inclusion of individuals who consider AI adoption non-strategic renders non-significative the relationship with the *training* dimension.

To better understand whether the assumption of not considering all the data is valid, it is necessary to go through the full uncertainty and SA as described to follow.

After comparing SEM estimates obtained via asymptotically inference and bootstrap and performing the counterfactual analysis, we implement here the third phase of our analysis. First, we propagate uncertainty in the SEM estimates through UA, and then we quantify the error propagation by providing total sensitivity indices of Sobol defined in Sect. 2.2.2.

In Equation (1) we added as a source of uncertainty random triggers, independent indicator variables ($\mathbb{1}$) that randomly assume a value equal to zero or to one for each item considered in the SEM, to determine whether the item is selected at any given simulation.

In this way, we define the latent variables as follows:

- The latent variable *AI adoption* is defined by $\{14_1 \mathbb{1}_{14_1}, 14_2 \mathbb{1}_{14_2}, 14_3 \mathbb{1}_{14_3}, 14_5 \mathbb{1}_{14_5}\}$
- The latent variable *vigour* is defined by $\{17_1 \mathbb{1}_{17_1}, 17_2 \mathbb{1}_{17_2}, 17_3 \mathbb{1}_{17_3}\}$
- The latent variable *dedication* is defined by $\{17_7 \mathbb{1}_{17_7}, 17_8 \mathbb{1}_{17_8}, 17_9 \mathbb{1}_{17_9}, 17_{10} \mathbb{1}_{17_{10}}\}$
- The latent variable *absorption* is defined by $\{17_{12} \mathbb{1}_{17_{12}}, 17_{14} \mathbb{1}_{17_{14}}, 17_{15} \mathbb{1}_{17_{15}}, 17_{16} \mathbb{1}_{17_{16}}\}$
- the latent variable *training* is defined by $\{20_2 \mathbb{1}_{20_2}, 20_3 \mathbb{1}_{20_3}, 20_5 \mathbb{1}_{20_5}, 20_6 \mathbb{1}_{20_6}\}$

At each iteration, we sample randomly different combinations of triggers in an independent way, that leads to an identifiable model. Moreover, we use a trigger $\mathbb{1}_{Data} = \{0, 1\}$ to test the data structural assumption used in Theben et al. (2023) to include in the analysis only individuals who consider the AI adoption strategic for their companies. If $\mathbb{1}_{Data} = 0$, we include only individuals who consider the AI

Table 5 Model estimates of Equation (1) provided by asymptotically inference and Uncertainty analysis with respectively their 95% percentage confidence and plausible intervals (* stands for a p-value < 0.05)

Coefficient	Asymptotically (range)	Uncertainty (plausible interval)
β_3^1	0.090 (−0.090–0.270)	0.129 (−0.120–0.383)
γ_1^1	0.807* (0.706–0.909)	0.784 (−0.810–0.912)
γ_2^1	0.408* (0.255–0.562)	0.392 (−0.469–0.582)
γ_3^1	−0.131 (−0.294–0.033)	−0.107 (−0.310–0.247)
ω_1^1	0.758* (0.677–0.839)	0.752* (0.611–0.906)
ω_2^1	0.436* (0.309–0.563)	0.437* (0.283–0.588)
ω_3^1	−0.185* (−0.321–−0.050)	−0.184 (−0.374–0.032)
α_1^1	0.794* (0.684–0.904)	0.817* (0.510–0.957)
α_2^1	0.325* (0.172–0.479)	0.302* (0.146–0.487)
α_3^1	−0.013 (−0.162–0.136)	−0.016 (−0.257–0.231)

adoption strategic for their companies, otherwise, we also include individuals who consider AI adoption to be non-strategic for their companies.

More precisely, we defined twenty triggers, one for each item that composed the latent variable (we do not need a trigger for the *complexity* dimension because it is composed of a single item), and one trigger for the selection of the data. In summary, we considered the following triggers:

$$\{\mathbb{1}_{14_1}, \mathbb{1}_{14_2}, \mathbb{1}_{14_3}, \mathbb{1}_{14_5}, \mathbb{1}_{17_1}, \mathbb{1}_{17_2}, \mathbb{1}_{17_3}, \mathbb{1}_{17_7}, \mathbb{1}_{17_8}, \mathbb{1}_{17_9}, \\ \mathbb{1}_{17_{10}}, \mathbb{1}_{17_{12}}, \mathbb{1}_{17_{14}}, \mathbb{1}_{17_{15}}, \mathbb{1}_{17_{16}}, \mathbb{1}_{20_2}, \mathbb{1}_{20_3}, \mathbb{1}_{20_5}, \mathbb{1}_{20_6}, \mathbb{1}_{Data}\}$$

For generating UA and GSA, we define a design matrix A and B of dimension $1,000 \times 20$ as discussed in Sect. 2.2.2.

In addition, for all these different model combinations, we repeatedly bootstrap data and we estimate the SEM. The following table compares the classical confidence interval with our confidence interval that includes these sources of uncertainty, and that, from here on, we call "plausible interval".

The results in Table 5 show that when a source of uncertainty in the model assumptions is considered, the plausible intervals are bigger than confidence intervals. In general, the central point estimates, given by the median of the distribution of the coefficients for the plausible intervals, remain stable. These results can be interpreted as the a confirmation of the model coefficient estimates provided by the original work of Theben et al. (2023). More precisely, the regression $Y_1^1 = \beta_3^1 X_3^1$ remains quite stable. In the regression $Y_2^1 = \gamma_1^1 X_1^1 + \gamma_2^1 X_2^1 + \gamma_3^1 X_3^1$, we note that the inclusion of a source of uncertainty makes the relation with γ_1^1, γ_2^1, and γ_3^1 non-significative. In the regression, $Y_3^1 = \omega_1^1 X_1^1 + \omega_2^1 X_2^1 + \omega_3^1 X_3^1$, the inclusion of a source of uncertainty makes the relation with ω_3^1 non-significative. The regression $Y_4^1 = \alpha_1^1 X_1^1 + \alpha_2^1 X_2^1 + \alpha_3^1 X_3^1$ remain stable. The point of this analysis is that, if the study investigated (Theben et al. (2023)) resists, as it does, to the modelling of the modelling process, this is a validation of the work done by Theben

Table 6 Total sensitivity indexes $S_{i,j}^{tot}$ for each trigger calculated on the parameters defined in Equation (1) estimated considering the source of uncertainty

Dimension		*AI adoption*				*complexity*			*training*		
Question	Trigger	β_3^1	γ_3^1	ω_3^1	α_3^1	γ_1^1	ω_1^1	α_1^1	γ_2^1	ω_2^1	α_2^1
AI adoption	$\mathbb{1}_{14\_1}$	0.14	0.34	0.37	0.45	0.09	0.09	0.01	0.25	0.26	0.22
	$\mathbb{1}_{14_2}$	0.15	0.26	0.18	0.29	0.10	0.08	0.01	0.14	0.14	0.07
	$\mathbb{1}_{14_3}$	0.17	0.35	0.20	0.32	0.07	0.04	0.01	0.22	0.20	0.14
	$\mathbb{1}_{14_5}$	0.36	0.19	0.15	0.21	0.03	0.04	0.01	0.10	0.15	0.03
Vigour	$\mathbb{1}_{17_1}$	0.00	0.07	0.00	0.01	0.23	0.02	0.00	0.04	0.00	0.00
	$\mathbb{1}_{17_2}$	0.00	0.05	0.00	0.00	0.03	0.02	0.00	0.05	0.00	0.00
	$\mathbb{1}_{17_3}$	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.01	0.00	0.00
Dedication	$\mathbb{1}_{17_7}$	0.00	0.00	0.02	0.00	0.02	0.05	0.00	0.00	0.01	0.00
	$\mathbb{1}_{17_8}$	0.01	0.01	0.11	0.02	0.13	0.36	0.02	0.00	0.05	0.01
	$\mathbb{1}_{17_9}$	0.01	0.01	0.09	0.01	0.08	0.07	0.01	0.01	0.13	0.01
	$\mathbb{1}_{17_{10}}$	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.00	0.02	0.00
Absorption	$\mathbb{1}_{17_{12}}$	0.00	0.01	0.01	0.19	0.05	0.08	0.06	0.00	0.00	0.03
	$\mathbb{1}_{17_{14}}$	0.01	0.01	0.01	0.22	0.14	0.14	0.53	0.01	0.01	0.07
	$\mathbb{1}_{17_{15}}$	0.00	0.00	0.00	0.06	0.03	0.03	0.03	0.00	0.00	0.03
	$\mathbb{1}_{17_{16}}$	0.00	0.00	0.00	0.04	0.01	0.01	0.01	0.00	0.00	0.02
Training	$\mathbb{1}_{20_2}$	0.02	0.04	0.02	0.06	0.21	0.20	0.04	0.26	0.30	0.32
	$\mathbb{1}_{20_3}$	0.04	0.10	0.06	0.15	0.31	0.26	0.07	0.41	0.36	0.53
	$\mathbb{1}_{20_5}$	0.03	0.07	0.04	0.11	0.29	0.27	0.09	0.27	0.34	0.37
	$\mathbb{1}_{20_6}$	0.01	0.01	0.01	0.01	0.10	0.13	0.01	0.22	0.34	0.18
Data	$\mathbb{1}_{Data}$	0.35	0.61	0.32	0.36	0.17	0.22	0.30	0.41	0.30	0.35

et al. (2023). This was not a foregone conclusions as shown by our discussion above of the results in Breznau, N., et al. (2022).

We calculated the total sensitivity index defined in Equation (3) for each of the coefficients estimated in SEM. Note that from the computational point of view, UA and uncertainty quantification are run simultaneously, so the results of the total index pertain to the same results discussed above.

The following table shows for each trigger total sensitivities index calculated on the parameters defined in Equation (1) estimated considering the source of uncertainty previously defined.

Note that our procedure of uncertainty propagation samples the triggers as independent from one another, which is an approximation but serve the purpose of testing robustness (Table 6).

As we can see from Table 6, in general, the inclusion of individuals who consider AI adoption to not be strategic inside the analysis has a large impact on the estimation of the coefficient.

Let us focus on the coefficient dependent on the *AI adoption* dimension:

- The estimation of β_3^1 is more influenced by the presence of item 14_5 in the definition of the dimension *AI adoption*;
- The estimation of γ_3^1 is more influenced by the presence of item 14_3 in the definition of the dimension *AI adoption*;
- The estimation of ω_3^1 is more influenced by the presence of item 14_1 in the definition of the dimension *AI adoption*;
- The estimation of α_3^1 is more influenced by the presence of item 14_1 in the definition of the dimension *AI adoption*.

The other item has a marginal effect in the estimation of regression coefficients dependent on the *AI adoption* dimension, including item 14_2, might not even be considered in the definition of the dimension.

Considering now the coefficient dependent on the *complexity* dimension:

- The estimation of γ_1^1 is more influenced by the presence of item 20_3 in the definition of the dimension *training*;
- The estimation of ω_1^1 is more influenced by the presence of item 17_8 in the definition of the dimension *dedication*;
- The estimation of α_1^1 is more influenced by the presence of item 17_{14} in the definition of the dimension *absorption*.

The other item has a marginal effect in the estimation of regression coefficients dependent on the *complexity* dimension.

Finally, let us give an interpretation of the coefficient dependent on the *training* dimension:

- The estimation of γ_2^1 is more influenced by the presence of item 20_3 in the definition of the dimension *training*;
- The estimation of ω_2^1 is more influenced by the presence of item 20_3 in the definition of the dimension *training*;
- The estimation of α_2^1 is more influenced by the presence of item 20_3 in the definition of the dimension *training*.

The other item (γ_2^1 or ω_2^1 or α_2^1) has a marginal effect in the estimation of regression coefficients dependent on the *training* dimension; including items 20_2, 20_5, and 20_6 might not even be considered in the definition of the dimension. Regular training and training updates represent the more important aspect in the definition of the latent variable *training*.

Test case 2

Following a similar procedure of *Test case 1*, as a first step of our analysis, we compared, in the following Table 7 and Fig. 4, model estimates of Equation (2) provided by asymptotically inference with the confidence intervals derived from bootstrap procedure repeated for one-thousand of times (Tibshirani and Efron (1993)).

Table 7 Model estimates of Equation (2) provided by asymptotically inference and bootstrap with their 95% percentage confidence intervals (* stands for a p-value < 0.05)

Coefficient	Asymptotically	Bootstrap
β_1^2	0.437* (0.252–0.622)	0.433* (0.264–0.661)
γ_1^2	0.688* (0.466–0.909)	0.691* (0.337–0.852)
γ_2^2	0.223* (0.006–0.441)	0.223* (0.006–0.509)
ω_1^2	0.594* (0.189–1.000)	0.560* (0.149–0.985)
ω_2^2	0.657* (0.382–0.931)	0.650* (0.363–0.967)
ω_3^2	−0.209 (−0.628–0.210)	−0.172 (−0.632–0.210)

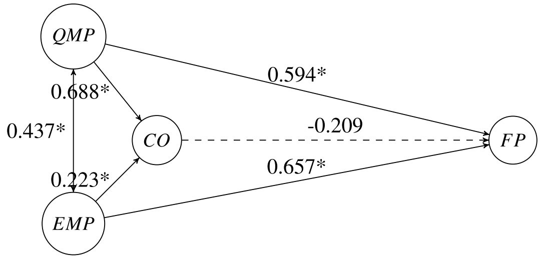

Fig. 4 Graphical representation of the SEM of Perramon et al. (2022). Above the arrows, we report the regression coefficient (* stands for a p-value < 0.05). Dashed arrow represent non-significative coefficients

Also in this case, table 7 shows that central points of model estimates of Equation (2), with their 95% confidence intervals provided by asymptotically inference, are coherent with the one obtained from bootstrap. In general, the dimensions *FP* are better explained by the dimensions *QMP* and *EMP* and the relationship with *CO* is non-significant. The dimension *CO* is explained by the dimension *QMP* and *EMP* and the two dimension *QMP* abd *EMP* are each other related.

We propagate uncertainty in the SEM estimates through UA, and then we quantify the error propagation by providing total sensitivity indices of Sobol defined in Sect. 2.2.2.

In Equation (2) we added as a source of uncertainty random triggers to determine whether the item is selected at any given simulation. Note that, at each iteration, we sample randomly different combinations of triggers in an independent way, that leads to an identifiable model.

In this way, we define the latent variables as follows:

- The latent variable QMP is defined by $\{QMP_1\mathbb{1}_{QMP_1}, QMP_2\mathbb{1}_{QMP_2}, QMP_3\mathbb{1}_{QMP_3}, QMP_4\mathbb{1}_{QMP_4}\}$;
- The latent variable EMP is defined by $\{EMP_1\mathbb{1}_{EMP_1}, EMP_2\mathbb{1}_{EMP_2}, EMP_3\mathbb{1}_{EMP_3}, EMP_4\mathbb{1}_{EMP_4}\}$;
- The latent variable CO is defined by $\{CO_1\mathbb{1}_{CO_1}, CO_2\mathbb{1}_{CO_2}, CO_3\mathbb{1}_{CO_3}, CO_4\mathbb{1}_{CO_4}\}$;
- The latent variable FP is defined by $\{FP_1\mathbb{1}_{FP_1}, FP_2\mathbb{1}_{FP_2}, FP_3\mathbb{1}_{FP_3}\}$,

and so we obtained fifteen triggers

Table 8 Model estimates of Equation (2) provided by asymptotically inference and bootstrap with their 95% percentage confidence intervals (* stands for a p-value < 0.05)

Coefficient	Asymptotically (range)	Uncertainty (plausible interval)
β_1^2	0.437* (0.252–0.622)	0.464* 0.242 0.730
γ_1^2	0.688* (0.466–0.909)	0.679* 0.297 0.923
γ_2^2	0.223* (0.006–0.441)	0.186 −0.123 0.516
ω_1^2	0.594* (0.189–1.000)	0.521* 0.029 1.214
ω_2^2	0.657* (0.382–0.931)	0.543* 0.087 0.980
ω_3^2	−0.209 (−0.628–0.210)	−0.064−0.712 0.378

$$\{\mathbb{1}_{QMP_1}, \mathbb{1}_{QMP_2}, \mathbb{1}_{QMP_3}, \mathbb{1}_{QMP_4}, \mathbb{1}_{EMP_1}, \mathbb{1}_{EMP_2}, \mathbb{1}_{EMP_3}, \mathbb{1}_{EMP_4},$$
$$\mathbb{1}_{CO_1}, \mathbb{1}_{CO_2}, \mathbb{1}_{CO_3}, \mathbb{1}_{CO_4}, \mathbb{1}_{FP_1}, \mathbb{1}_{FP_2}, \mathbb{1}_{FP_3}\}.$$

For generating UA and GSA, we define a design matrix A and B of dimension $1,000 \times 15$ as discussed in Sect. 2.2.2. For all these different model combinations, we repeatedly bootstrap data and we estimate the SEM. The following table compares the classical confidence interval with our confidence interval that includes these sources of uncertainty, and that, from here on, we call “plausible interval".

Also in this test case, the results in Table 8 show that when a source of uncertainty in the model assumptions is considered, the plausible intervals are bigger than confidence intervals but, in general, the results remain stable. Only in the regression $Y_2^2 = \gamma_1^2 X_1^2 + \gamma_2^2 X_2^2$, we note that the inclusion of a source of uncertainty makes the relation with γ_2^2 non-significative.

We calculated the total sensitivity index defined in Equation (3) for each of the coefficients estimated in SEM.

The following table shows for each trigger total sensitivities index calculated on the parameters defined in Equation (2) estimated considering the source of uncertainty previously defined.

Let us focus on the coefficient dependent on the *QMP* dimension of Table 9:

- The estimation of β_1^2 is more influenced by the presence of item EMP_3 in the definition of the dimension *EMP*;
- The estimation of γ_2^1 is more influenced by the presence of item QMP_2 in the definition of the dimension *QMP*;
- The estimation of ω_1^2 is more influenced by the presence of item EMP_1 in the definition of the dimension *EMP*.

The other item has a marginal effect in the estimation of regression coefficients dependent on the *QMP*.

Considering now the coefficient dependent on the *EMP* dimension:

Table 9 Total sensitivity indexes $S_{i,j}^{tot}$ for each trigger calculated on the parameters defined in Equation (1) estimated considering the source of uncertainty

Dimension		*QMP*			*EMP*		*CO*
Question	Trigger	β_1^2	γ_1^2	ω_1^2	γ_2^2	ω_2^2	ω_3^2
QMP	$\mathbb{1}_{QMP_1}$	0.13	0.32	0.19	0.16	0.28	0.03
	$\mathbb{1}_{QMP_2}$	0.21	0.38	0.16	0.19	0.28	0.02
	$\mathbb{1}_{QMP_3}$	0.15	0.31	0.15	0.17	0.21	0.01
	$\mathbb{1}_{QMP_4}$	0.20	0.34	0.18	0.20	0.27	0.03
EMP	$\mathbb{1}_{EMP_1}$	0.29	0.14	0.26	0.13	0.03	0.13
	$\mathbb{1}_{EMP_2}$	0.07	0.03	0.03	0.09	0.04	0.18
	$\mathbb{1}_{EMP_3}$	0.53	0.14	0.12	0.11	0.11	0.30
	$\mathbb{1}_{EMP_4}$	0.21	0.14	0.14	0.23	0.15	0.53
CO	$\mathbb{1}_{CO_1}$	0.01	0.15	0.15	0.20	0.06	0.02
	$\mathbb{1}_{CO_2}$	0.00	0.15	0.23	0.11	0.05	0.01
	$\mathbb{1}_{CO_3}$	0.00	0.10	0.08	0.09	0.04	0.02
	$\mathbb{1}_{CO_4}$	0.01	0.22	0.13	0.08	0.03	0.02
FP	$\mathbb{1}_{FP_1}$	0.03	0.01	0.01	0.22	0.31	0.15
	$\mathbb{1}_{FP_2}$	0.23	0.06	0.01	0.07	0.05	0.04
	$\mathbb{1}_{FP_3}$	0.02	0.02	0.01	0.02	0.04	0.02

- The estimation of γ_2^2 is more influenced by the presence of item EMP_4 in the definition of the dimension *EMP*;
- The estimation of ω_2^2 is more influenced by the presence of item QMP_1 and QMP_2 in the definition of the dimension *QMP*.

The other item has a marginal effect in the estimation of regression coefficients dependent on the *EMP* dimension.

Finally, let us give an interpretation of the coefficient dependent on the *CO* dimension, where the estimation of γ_3^2 is more influenced by the presence of item EMP_4 in the definition of the dimension *EMP*.

4 Discussion and conclusion

SEM proposed by Bentler (1983), is a widely used multivariate technique to test relations among observed and latent variables. In the literature, there are two main types of SEM (Dash and Paul (2021)): CB-SEM, for a factor-based model, and PLS-SEM, for a composite-based model. The test cases here presented (Theben et al. (2023); Perramon et al. (2022)) used CB-SEM. This choice was maintained in the present analysis, which presents a new approach to provide SA on SEM when perturbations in the composition of latent variables are introduced.

SA is a general statistical concept to evaluate the stability of estimators concerning parameters and model assumptions, as discussed in Saltelli et al. (2008) and Leamer (1984). In this paper, the method proposed by Saltelli et al. (2000) is applied to robustify SEM estimates and detect the influential subsets of data or variables that seriously influence the analysis. This use of sensitivity analysis to explore the construction of a model has also been called “modelling of the modelling process" (Piano et al. (2022); Saltelli and Puy (2023)).

Quantifying the impact of controlled perturbations to modeling conditions on study results is an important diagnostic tool given its direct implications on the inferences, especially in light of the present discussions on the possible fragility of model-based inference to modeling assumptions, as discussed in, e.g., Gelman and Loken (2013) and in the experiment Breznau, N., et al. (2022).

As a first step of the analysis, we tested if the bootstrap procedure is coherent with the original results of the two test cases. After this step, we added triggers as a source of uncertainty while simultaneously bootstrapping the data to insert sample variability in the model estimates. Finally, through GSA, we quantified the error propagation by providing total sensitivity indices defined in Sect. 2.2.2 in Equation (3). The total indices provided by GSA gives us a measure defined in [0, 1] to better understand the impact of a single item on the estimates of the coefficients of the regression. To note that Sobol indices here presented (Saltelli et al. (2008)) are valid only under the assumption of input independence, although approaches that address this limitation are available in the literature (Mara et al. (2015)).

This study corroborates and robustifies previous findings. In general GSA substantially confirms the previous research taken as our test cases. More precisely, in

- Theben et al. (2023) we found that:
  - *AI adoption* has a negative and in general non-significant relationship with employee engagement and the *complexity* dimension appears to have a positive and significant relationship with these dimensions;
  - *Vigour* could have a non-significative association with *training* and *compelxity* and *AI adoption* could have a non-significative association with *dedication*.
- Perramon et al. (2022) we found that:
  - *EMP* has a direct correlation with other key practices of the company, *EMP* and *QMP* are relevant to obtain higher *CO* and *FP*, and non-significant relationship between *CO* and *FP*;
  - *CO* could have a non-significative association with *EMP*.

In Theben et al. (2023), we note that including or not including individuals who think that AI adoption is not strategic for their companies makes a difference in the results; thus, Theben et al. (2023) were right to limit the sample to respondents who consider AI adoption strategic. In booth test case (Theben et al. (2023); Perramon et al. (2022)), total index of each item provided by GSA shows that the dimension created via factor analysis is robust to model perturbations

However, the present paper has limitations worth discussing. The first is that, in these test cases, we did not consider the scenario of all manifest variables of a latent variable are associated with a trigger that takes the value 0. We considered this scenario rare and we preferred not to change the graphical structure of the model.

Although we consider it irrelevant to these case studies, it is possible to introduce this case study in future developments. A second limitations is represented by the data used in Theben et al. (2023). The data were collected through a self-reported cross-sectional survey, which does not allow for causal conclusions to be drawn. For this reason, we introduce a second test case (Perramon et al. (2022)) that contains not-self-reported cross-sectional data. In addtion, our intention is not to provide causal interpretation of the phenomenon, but only to demonstrate the feasibility and usefulness of a robustification analysis applied to SEM. In-fact, thanks to GSA, we are able to investigate the equifinality problem, in which multiple model structures are compatible with the same data (Morgan and Morrison (1999)).

Since our analysis was entirely based on open-source software, the procedure described here is easy to apply to cases where a similar methodology is employed. The GSA represents an optimal tool for validating the assumptions made by the modeler. Stopping the analysis at the estimation phase, turns out to be an incorrect procedure because the statistical inference is done with respect to a functional form of the model. It is important to reiterate the fact that all modelling choices done by modellers are explicit and subject to a process of estimation.

Another important application of GSA, not explored in this paper, is its utility in identifying the best-fitting model. Becker et al. (2021) demonstrated that GSA serves as an effective tool for variable selection in regression models, using model fit as the output measure. This approach could, of course, be extended to SEM.

In this application, we have for example varied the data used for the estimation procedure of SEM, we have created different definitions of latent variables, which are often a concept extremely subjective of the modeller, and we have included the phase of bootstrap of data. All these variations, which we could call "model assumptions", represents a typical case of something which might be assumed as inconsequential and that instead might affect the results. Future work, for example, could introduce a comparison via GSA of the CB-SEM versus PLS-SEM approach, as already discussed by Dash and Paul (2021) and Astrachan et al. (2014).

Author contributions A.S. and A.L. conceptualized the project; A.L. wrote the codes and conducted the statistical analysis under the supervision of A.S.; A.L. and A.S. wrote the first version of the paper; J.L. and J.P. wrote the paper; M.B. provided critical feedback. All the authors read and approved the final version of the paper.

Data availibility statement Data are available on request from the corresponding author.

Code availability Codes are available on request from the corresponding author.

Declarations

Conflict of interest The authors declared no potential Conflict of interest with respect to the research, authorship, and/or publication of this article.

References

Astrachan CB, Patel VK, Wanzenried G (2014) A comparative study of cb-sem and pls-sem for theory development in family firm research. J Family Bus Strat 5(1):116–128

Bagur-Femenías L, Perramon J, Oliveras-Villanueva M (2019) Effects of service quality policies in the tourism sector performance: an empirical analysis of spanish hotels and restaurants. Sustainability 11(3):872

Ballester-Ripoll R, Leonelli M (2021) Global sensitivity analysis in probabilistic graphical models. arXiv preprint arXiv:2110.03749

Barr J, Rabitz H (2022) A generalized kernel method for global sensitivity analysis. SIAM/ASA J Uncertain Quant 10(1):27–54

Barr J, Rabitz H (2023) Kernel-based global sensitivity analysis obtained from a single data set. Reliab Eng Syst Saf 235:109173

Becker W, Paruolo P, Saltelli A (2021) Variable selection in regression models using global sensitivity analysis. J Time Ser Econ 13(2):187–233

Bentler PM (1983) Some contributions to efficient statistics in structural models: specification and estimation of moment structures. Psychometrika 48:493–517

Bentler P, Wu E (2015) Supplement to EQS 6.3 for Windows User's Guide

Borges JL (1998) The garden of forking paths, Penguin books edn., London. https://www.penguin.co.uk/books/308559/the-garden-of-forking-paths-by-borges-jorge-luis/9780241339053

Borgonovo E (2007) A new uncertainty importance measure. Reliab Eng Syst Saf 92(6):771–784

Breznau N et al (2022) Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proceed Natl Acad Sci 119(44):e2203150119. https://doi.org/10.1073/pnas.2203150119

Cantone GG, Tomaselli V (2024) Characterisation and calibration of multiversal methods. Adv Data Anal Classif, 1–33

Chatfield C (1995) Model uncertainty, data mining and statistical inference. J Royal Stat Soc Ser A: Stat Soc 158(3):419–444

Dash G, Paul J (2021) Cb-sem vs pls-sem methods for research in social sciences and technology forecasting. Technol Forecast Social Change 173:121092

De Bock J, De Campos CP, Antonucci A (2014) Global sensitivity analysis for map inference in graphical models. Adv Neural Inform Process Syst **27**

Engzell P (2023) A universe of uncertainty hiding in plain sight. Proceed Natl Acad Sci 120(2):2218530120. https://doi.org/10.1073/pnas.2218530120

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 4(3):272

Finch JF, West SG (1997) The investigation of personality structure: statistical models. J Res Personal 31(4):439–485

Fontaine JR (2005) Equivalence. Encyclop Social Meas 1:803–813

Gelman A, Loken E (2013) The garden of forking paths: why multiple comparisons can be a problem even when there is no fishing expedition or p-hacking" and the research hypothesis was posited ahead of time. Dep Stat 348:3

Kaplan D (2008) Structural equation modeling: foundations and extensions vol 10. SAGE publications

Kroese DP, Brereton T, Taimre T, Botev ZI (2014) Why the monte carlo method is so important today. Wiley Interdiscip Rev: Comput Stat 6(6):386–392

Kucherenko S, Albrecht D, Saltelli A (2015) Exploring multi-dimensional spaces: a comparison of latin hypercube and quasi monte carlo sampling techniques. arXiv preprint arXiv:1505.02350

Leamer EE (1984) Global sensitivity results for generalized least squares estimates. J Am Stat Assoc 79(388):867–870

Lee S-Y, Wang S-J (1996) Sensitivity analysis of structural equation models. Psychometrika 61:93–108

Leite WL, Shen Z, Marcoulides K, Fisk CL, Harring J (2022) Using ant colony optimization for sensitivity analysis in structural equation modeling. Struct Equa Model: A Multidiscip J 29(1):47–56

Likert R (1932) A technique for the measurement of attitudes. Arch Psychol

Lomax RG (1983) A guide to multiple-sample structural equation modeling. Behav Res Methods Instrum 15(6):580–584

Mara TA, Tarantola S, Annoni P (2015) Non-parametric methods for global sensitivity analysis of model output with dependent inputs. Environ Modell softw 72:173–183

Morgan MS, Morrison M (eds) (1999) Models as mediators: perspectives on natural and social science, 1st edn. Cambridge University Press . https://doi.org/10.1017/CBO9780511660108 . https://www.cambridge.org/core/product/identifier/9780511660108/type/book Accessed 2023-09-15

Nachtigall C, Kroehne U, Funke F, Steyer R (2003) Pros and cons of structural equation modeling. Methods Psychol Res Online 8(2):1–22

Norris M, Lecavalier L (2010) Evaluating the use of exploratory factor analysis in developmental disability psychological research. J Autism Develop Disorders 40:8–20

Owen AB (2014) Sobol'indices and shapley value. SIAM/ASA J Uncertain Quant 2(1):245–251

Pek J, MacCallum RC (2011) Sensitivity analysis in structural equation models: cases and their influence. Multiv Behav Res 46(2):202–228

Perramon J, Oliveras-Villanueva M, Llach J (2022) Impact of service quality and environmental practices on hotel companies: an empirical approach. Int J Hosp Manag 107:103307

Piano SL, Lőrincz MJ, Puy A, Pye S, Saltelli A, Smith ST, Sluijs JV (2023) Unpacking the modeling process for energy policy making. Risk Analy. https://doi.org/10.1111/risa.14248

Piano SL, Sheikholeslami R, Puy A, Saltelli A (2022) Unpacking the modeling process via sensitivity auditing. Futures 144:103041

Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, Wagener T (2016) Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ Modell Softw 79:214–232

Pianosi F, Wagener T (2015) A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ Modell Softw 67:1–11

Razavi S, Gupta HV (2016) A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1 Theory. Water Resour Res 52(1):423–439

Razavi S, Jakeman A, Saltelli A, Prieur C, Iooss B, Borgonovo E, Plischke E, Piano SL, Iwanaga T, Becker W (2021) The future of sensitivity analysis: an essential discipline for systems modeling and policy support. Environ Modell Softw 137:104954

Ringle CM, Sarstedt M, Mitchell R, Gudergan SP (2020) Partial least squares structural equation modeling in HRM research. Int J Human Resour Manag 31(12):1617–1643

Saltelli A (2002) Sensitivity analysis for importance assessment. Risk Anal 22(3):579–590

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Phys Commun 181(2):259–270

Saltelli A, Puy A (2023) What can mathematical modelling contribute to a sociology of quantification? Human Social Sci Commun. https://doi.org/10.2139/ssrn.4212453

Saltelli A, Di Fiore M (2023) The politics of modelling: numbers between science and policy. Oxford University Press

Saltelli A, Tarantola S, Campolongo F (2000) Sensitivity analysis as an ingredient of modeling. Stat Sci, 377–395

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the Primer. John Wiley & Sons

Schuberth F, Rosseel Y, Rönkkö M, Trinchera L, Kline RB, Henseler J (2023) Structural parameters under partial least squares and covariance-based structural equation modeling: a comment on yuan and deng (2021). Struct Equa Model: A Multidiscip J 30(3):339–345

Sobol IM (1994) A primer for the Monte Carlo Method. CRC press

Sobol' IM (1993) Sensitivity analysis for non-linear mathematical models, mathematical modelling and computational experiment (Translated from Russian: I.M. Sobol', Sensitivity estimates for nonlinear mathematical models, Matematicheskoe Modelirovanie 2 (1990) 112-118) **1**(4), 407–414

Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W (2016) Increasing transparency through a multiverse analysis. Perspect Psycholog Sci 11(5):702–712

Theben A, Plamenova N, Perramon J, Llach J (2023) The impact of AI adoption on employee engagement: preparing the workforce for new realities. UPF-BSM working paper. https://cms.bsm.upf.edu/sites/default/files/inline-files/2023-manuscript-ai-adoption-engagement.pdf

Tibshirani RJ, Efron B (1993) An introduction to the Bootstrap vol 57

Wang S-J, Lee S-Y (1996) Sensitivity analysis of structural equation models with equality functional constraints. Comput Stat Data Anal 23(2):239–256

Wolfle LM (1982) Pase: program for analysis of structural equations. Behav Res Methods Instrum 14(6):548–550

Wolfle LM, Ethington CA (1985) Gemini: program for analysis of structural equations with standard errors of indirect effects. Behav Res Methods Instrum Computers

Authors and Affiliations

Alessio Lachi[1] · Josep Llach[3] · Jordi Perramon[3] · Michela Baccini[2] · Andrea Saltelli[3]

✉ Alessio Lachi
alessio.lachi@unicamillus.org

Josep Llach
Josep.Llach@bsm.upf.edu

Jordi Perramon
jordi.perramon@bsm.upf.edu

Michela Baccini
michela.baccini@unifi.it

Andrea Saltelli
andrea.saltelli@bsm.upf.edu

1 Department of Medicine, Saint Camillus International University of Health and Medical Sciences, Via di Sant'Alessandro 8, 00131 Roma, Italy

2 Department of Statistics, Computer Science, Applications "Giuseppe Parenti" (DiSIA), University of Florence, Viale Giovanni Battista Morgagni 59/65, 50134 Florence, Italy

3 UPF Barcelona School of Management (UPF-BSM), University Pompeu Fabra of Barcelona, Carrer de Balmes 132-134, 08008 Barcelona, Spain