

USING STRUCTURAL EQUATION AND ITEM RESPONSE MODELS TO ASSESS RELATIONSHIP BETWEEN LATENT TRAITS

Anna SIMONETTO¹

PhD, Researcher,
University of Brescia, Italy

E-mail: simonett@eco.unibs.it

Abstract: We deepen the two main approaches to the problem of measurement error in social sciences, the Structural Equation Models (SEM) and the Item Response Theory Models (IRM), comparing two different estimation procedures.

The One-step procedure (related to SEM) requires that researcher specifies a complete model of both measurement aspects (single link between the latent variable and its indicators) and structural aspects (links between different latent variables), with the model parameters estimated simultaneously. In the Two-step procedure (related to IRM), we first estimate the measures (one for each construct), then we will assess, through a regression model, the relationships between these measures and the latent variables that they represent.

Our aim is to define a Two-step method that, using information obtained in the first step about the measurement error, presents low levels of bias and loss of efficiency, as close as possible to that of One-step method.

Key words: latent variable models; structural equation models; item response theory; measurement error

1. Introduction

The starting point for this research is a concrete problem: to measure, using statistical models, subjective perceptions and assessments and to understand their dependencies. The purpose is to evaluate two different estimation procedures for regression models with variables affected by measurement errors. The first procedure, named One-step, considers simultaneously all the parameters involved in the complete Structural Equation Model (SEM) for the hypothesized latent and observed variables. The Two-step procedure starts obtaining, using an adequate Item Response Theory Model (IRM), the measures associated to latent variables. Then we derive parameter estimates of the latent variable regression model; in its specification we use the measures obtained at the first step, considering that they are affected by measurement errors.

We want to define a Two-step method that present low levels of bias and loss of efficiency, as close as possible to that of One-step method. In the simulation study, we will

evaluate the impact of the measurement error in the case of standard regression. Another original aspect of the work concerns the reliability index used to estimate the variance of measurement error: the Rasch Person Reliability Index.

In the next section, we frame the problem of categorical data analysis and we review the essential features of the principal methods in literature. In the third section, we present the two estimation procedures. In the fourth section, we describe the simulation conducted to study some estimator features and we report the results of applying the presented methods to a real case of job satisfaction analysis. In the final section, we summarize some considerations arising from the comparison of the proposed estimation procedures.

2. Categorical data analysis and measurement

When we have to analyze multidimensional aspects that are not directly observable or measurable through traditional survey instruments, we want to define a scientific measurement for them, taking into account all the existing links between the different aspects involved. These concepts are defined as constructs, latent factor or latent variables, that are not directly observable but that can be inferred, through a mathematical model, from other variables that we can observe and directly measure.

The scientific measurement in economic and social sciences would match the same standards of scientific measurement in the physical sciences and the goal of researchers is to determine the most reproducible and additive measures that are objective abstractions of equal units. To summarize, we can identify two main approaches to analyzing multivariate latent aspects taking into account the categorical nature of the observed variables (Cagnone, Mignani, and Moustaki, 2010): the Underlying Variable Approach (UVA) and the Item Response Theory (IRT).

The UVA assumes that the observed categorical outcomes are incomplete observations of unobserved continuous variables. Underlying each of the categorically observed variables there is a continuous variable that measures the underlying latent factor, not directly observable. To fit this model have been proposed several methods, we will refer to the model of Muthén (1984), implemented in the *Mplus* software.

With the IRT approach, the unit of analysis is the entire response pattern of a subject, so we have no loss of information. For a given observed variable, we can write its distribution as a function of the latent trait level.

We will compare the results obtained by applying an UVA and an IRT model (SEM and IRM respectively) to the same datasets. The idea of how to make this comparison has been taken by a work of Gibbons et al. (2007): the authors indirectly showed, proposing their bifactorial model, how to implement an IRM in a SEM framework.

2.1. The Underlying Variable Approach and the Structural Equation Model

In the context of the UVA, the SEM had a very wide spread. We are interested in the specification and estimation for type of SEM with latent variables having multiple indicators. For the continuous latent variables we consider a linear structure, while, in the measurement part, we could have dichotomous, ordered polytomous and/or continuous observed indicators.

The model used in the simulation study are drawn from several works of Muthén (for references see Muthén and Muthén, 2007). He makes a distinction between models

with observed independent variables and model without them, but we consider only the last ones because they are more closely to IRM that we have implemented in our study.

A full SEM can be split in two submodels:

- The *Structural Model* refers to the latent variables in the model and it expresses the hypothesized relationships among the constructs

$$\theta = B\theta + \zeta;$$

- the *Measurement Model* links the constructs to observable indicators

$$y^* = \Lambda^T \theta + \epsilon.$$

We consider the $(m \times 1)$ vector θ of continuous latent variable. B is the $(m \times m)$ matrix of the coefficients for the effects of each variable on each other; it has zero diagonal elements and $(I - B)$ is non-singular. ζ is the $(m \times 1)$ vector of disturbances, and it represents the structural equation errors. y^* is the $(p \times 1)$ vector of continuous latent response variables associated with the observed variables. Λ^T is the $(p \times m)$ matrix of coefficients (loadings) about the relations of y^* with θ . The $(p \times 1)$ vector $\epsilon \sim N(0, I)$ represents the measurement errors. Regression models implicitly assume the absence of measurement error and so, if such error exists, regression coefficients are attenuated. In SEM the error terms are explicitly modeled, so SEM estimators are unbiased by error terms.

For an ordered polytomous y_j with c_j categories, we have

$$y_j = \begin{cases} c_j - 1, & \text{if } \tau_{j,c_j-1} < y_j^* \\ c_j - 2, & \text{if } \tau_{j,c_j-2} < y_j^* \leq \tau_{j,c_j-1} \\ \vdots & \\ 1, & \text{if } \tau_{j,1} < y_j^* \leq \tau_{j,2} \\ 0, & \text{if } y_j^* \leq \tau_{j,1} \end{cases} \quad (1)$$

2.2. Item Response Theory and Item Response Model

In the IRT approach, the purpose is to obtain an objective measure of the latent construct of interest. Within this framework, several models have been proposed to produce an objective measure of the latent construct, synthesizing data obtained from a questionnaire. The goal of an Item Response Theory Model (IRM) is to describe, through a nonlinear monotonic function, the association between a respondent's underlying latent trait level and the probability of a particular item response (Furr and Bacharach, 2008). To check whether the data fit satisfactorily to the model, it is possible to use some diagnostic tools based on the calculation of the residuals.

In an IRM we find two kinds of parameters, one that describes the qualities of the subject under investigation (ability), and the other relates to the characteristics of each item (difficulty) (Hays et al., 2000). The incorporation of linear structures allows for modeling the effects of covariates and enables the analysis of repeated categorical measurements.

When we have polytomous responses, two of the more widely used models are the Partial Credit Models (PCM - Masters, 1982) and Rating Scale Model (RSM - Andrich, 1978). Consistent with that proposed by Gibbons et al (2007), we chose the RSM. The two main features of this model are that items have the same number of categories (c) and the difference between any given threshold location and the mean of the threshold locations is equal or uniform across items. Furthermore, in the RSM all the p items are assumed to provide the same amount of information.

For the simulation study, we adopted the formulation of the RSM, belonging to the very general family of the Extended Rasch Models, proposed by Mair and Hatzinger (2007) and implemented in the R package "eRm". The RSM probability for the i -th subject to answer at the j -th item the k -th category is:

$$P(y_{ij} = k) = \frac{\exp[k(\theta_i + \delta_j) + \omega_k]}{\sum_{l=0}^{k-1} \exp[lk(\theta_i + \delta_j) + \omega_l]} \quad (2)$$

where ω_k are the category parameters ($\omega_k = \sum_{l=0}^k \tau_l$). The location parameter δ_j represents the average difficulty for a particular item relative to the category intersections. θ_i identifies the level of latent aspect possessed by the subject i , while the threshold parameter τ_l (with $\tau_0 = 0$ and $\sum_{l=0}^k \tau_l = 0$) quantifies the difficulty of choosing the l -th answer rather than the previous one.

θ_i , δ_j and τ_l are measured on a logit scale (log odds unit), so we can order them either by subjects (from the most satisfied to least satisfied) and by applications (from easiest to hardest). Furthermore, since both parameters are expressed in the same unit, it is possible to make cross-comparisons between subjects and questions.

2.3. Reliability analysis

When we use an IRM, we should evaluate of the reliability of the obtained measures. As usual, we assume that measures and errors are uncorrelated and that

$$\text{True variance } (S_x^2) = \text{Observed variance } (S_o^2) + \text{Error variance } (S_e^2),$$

so we can define the reliability of a measure as the proportion of its variance (the observed variance of the Rasch measure) attributable to the variance of the real underlying factor that we are measuring:

$$\text{Reliability} = \frac{\text{True variance}}{\text{Observed variance}} = \frac{S_x^2}{S_o^2}$$

Cronbach's alpha (α), thanks to its computational simplicity and easiness of understanding, is probably the most famous and popular reliability index (Cronbach, 1951). It has a general formula (DeVellis, 1991) from which derive many other indices (for example the Kuder-Richardson, KR, coefficients):

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\sum_{j=1}^p \sigma_j^2}{\sigma_t^2} \right).$$

p is the number of items; σ_t^2 is the variance of r (the observed raw scores for the current sample of persons) and σ_j^2 is the variance of the j -th item for the current sample of persons. Cronbach's α describes the internal consistency of groupings of items; an high value of this index indicates that the respondents express a coherent position on each item belonging to the same dimension.

In Rasch measurement we can use the person separation index instead of classical reliability indices. Rasch Person Reliability index (RPRI) (Linacre, 1997; Schumacker and Smith, 2007) indicates the replicability of person ordering that we could expect if another parallel set of items measuring the same construct were given at the same set of persons. This index is a ratio between the latent construct variance S_x^2 and the measure variance S_o^2 :

$$\text{RPRI} = \frac{S_x^2}{S_o^2} = 1 - \frac{S_e^2}{S_o^2}, \quad (3)$$

where $S_e^2 = S_o^2 - S_p^2$. Indicating the standard error for the i -th subject's ability estimate with $S_{e_i}^2$, we have

$$S_e^2 = \frac{1}{n} \sum_{i=1}^n S_{e_i}^2. \quad (4)$$

In the following paragraphs, the error variance of all measures obtained with IRM is estimated from the sum of the modeled variance of observations. This model error variance requires the data to conform stochastically to the proposed model. Rasch models provide a direct estimate of the model standard error for the estimate of a subject's ability ($S_{e_i}^2$), that gives a quantification of the precision of every person measure.

The relationship between raw-score-based reliability index (α) and measure-based reliability index (RPRI) is complex (Schumacker and Smith, 2007); in general, α overestimates reliability, RPRI underestimates it.

3. The estimation procedures

The purpose of this study is to compare, on the same data, the results obtained using two different estimation procedures, based on SEM and IRM respectively.

The two main research interests for the analysis of latent variable models with psychological traits are obtaining good measures and assessing the dependence relationships between the constructs they represent. Measures and dependence links may be combined into a single estimation procedure or developed sequentially one at a time.

- We have the **One-step procedure** when we combine the two interests in a single model. This procedure requires that the researcher specifies a complete model of both measurement aspects (single link between the latent variable and its indicators) and structural aspects (links between different latent variables). The model parameters are estimated simultaneously.
- We have the **Two-step procedure** when we estimate the measures and their dependence in two different phases. In the first step, we separately estimate the measures (one for each construct); in the second step we will assess, through a regression model, the relationships between these measures (and between the latent variables that they represent).

The One-step procedure should be more efficient, since it provides simultaneous estimation of latent variables and their dependence relationships. However, it does not allow to analyze the obtained measures in IRT perspective, that is the strength of Two-step procedure. Crucial element is to find a correct model that considers the measures obtained by the first step as affected by measurement errors.

We have implemented an articulated simulation study to evaluate the impact of this measurement error in the case of standard regression and it assesses whether the Two-step procedure is preferable compared to the One-step procedure. For comparison, we will consider the loss of efficiency and accuracy of the Two-step procedure, but we will evaluate which procedure allows better control in both phases: measures construction and regression model.

3.1. The one-step procedure

This procedure is combined with the UVA approach. The starting point are the latent variables underlying the observed responses and the relationships between these constructs. For this reason, the One-step procedure involves the simultaneous estimation of all model parameters through the implementation of the Muthén SEM, implemented in *Mplus* (Muthén and Muthén, 2007).

We implemented two different estimation methods for two different models:

- Structural Equation Model standard (**SEMstd**),
- Structural Equation Model based on the IRT approach (**SEMirt**).

SEMstd is the simplest model in this study. We are interested in the estimation of the regression coefficients and threshold parameters in (1). To homogenize the comparisons with the results obtained through other estimation methods and with different experimental conditions, we always standardize all the estimated variables: in other terms, we consider the " β coefficients".

SEMirt is a modified version of the previous model, inspired by the work of Gibbons et al. (2007), that introduces the structure of IRM in SEM. *Mplus* does not allow to specify directly the difficulty item parameters δ_j , so we have to introduce p fake latent variables Δ_j (one for each ordinal indicator item) which are formally latent variables, but their variance is imposed equal to 0. The Δ_j are completely uncorrelated with each other and with all other variables in the model. The means of these fake variables represent the difficulty item parameters δ_j . Furthermore, to recreate IRM, we have to impose that all relevant loadings be equal to 1. So, using the notation in section 2.1, we have:

The Structural Model

$$\theta = B\theta + \zeta \quad (5)$$

$$\Delta = \delta + \zeta_{\Delta}. \quad (6)$$

Δ is the vector ($p \times 1$) of fake variables, we impose $\text{var}(\zeta_{\Delta} \equiv 0)$, and its mean represents the difficulty of the p items.

The Measurement Model

$$y^* = \Lambda^y \theta + \Delta + \epsilon, \quad (7)$$

where Λ^y is the ($p \times m$) loading matrix. To lead us back to the IRM structure λ_{ik} is imposed equal to 1 if the k -th item refers to the l -th construct, otherwise $\lambda_{ik} = 0$.

With this model, we can estimate the β coefficients, the threshold parameters in (1) and, in addition to the previous model, the item difficulty parameters δ . It is important to underlie that all the model error terms are considered to be uncorrelated with each other and with other variables in the model. The variance of the structural errors ζ_h will be indicated with ψ_h .

Several indices of goodness of fit have been proposed for SEM, but it is not possible to proceed to the reliability analysis and all the other considerations (for example on the unidimensionality of constructs, the item analysis and correct categories order) that represent a significant part of the Two-step procedure.

3.2. The two-step procedure

In this procedure we combine the IRT approach, which focuses on observed variables, with the measurement error models. In the first step we estimate, through a IRM, the measure of each latent variable. These measures are then entered into a regression model to estimate the dependence relationships between the constructs.

To respect unidimensionality condition required by IRM, we divide the items according to the latent trait they refer to. Once the various measures are constructed, for each of them we can analyze the goodness of results. The IRM includes an analysis phase where researchers have to determine if the constructed measure meets all the main features of the model (for example unidimensionality, category proper order, reliability).

Once this analysis successfully, in the second step we want to estimate the dependence relationships between the constructs. We implement a linear regression model where we use the measures obtained in the first step as regressors and response variables. Applying this procedure, we should consider that the measures are affected by measurement error, estimated through (4), that can greatly influence the estimation of parameters in the second step.

We define two different models for this procedure, with the intention of being able to evaluate the essential characteristics of the estimation methods on simulated data:

- Rating Scale Model - Linear Regression Model with Measurement Error (**RSM-LRMme**);
- Rating Scale Model - Standard Linear Regression Model (**RSM-LRM**).

Crucial element is to find a correct model that considers the measures obtained in the first step as measures affected by measurement errors. **RSM-LRMme** has been implemented to obtain estimates with low levels of bias and loss of efficiency, as close as possible to those of One-step methods. The second model ignores that the measures are affected by measurement error.

The first step of both methods is the same: for each of the m constructs, we estimate its measure, Y^* , through an extended RSM (2), and their person reliability, using the standard errors of the person parameters (Mair and Hatzinger, 2007). Before moving to the next step, we standardize the estimated measures and all quantities involved in the model to obtain the beta weights, β , comparable with those obtained by the other estimation methods or with different experimental conditions.

The second step changes between the two methods. To explain the differences, we assume to have two measures, Y_1^* and Y_2^* , (obtained in the previous step), which are a function of the constructs, θ_1 and θ_2 , plus measurement errors, ϵ_1 and ϵ_2 , respectively.

In **RSM-LRMme**, we implemented a linear regression model taking into account that the model variables are affected by measurement errors. It is important to include this information in the model, to compensate for the attenuation effect due to measurement error. We used the Fox (2006) approach to SEM, implemented in the R package "sem", that refers to the Reticular Action Model. For simplicity, we use only a subscript for the coefficients, so the regression equation is:

$$\theta_1 = \beta_2 \theta_2 + \zeta_1.$$

and the measurement error equations are

$$Y_1^* = \lambda_1 \theta_1 + \epsilon_1$$

$$Y_2^* = \lambda_2 \theta_2 + \epsilon_2.$$

As measurement error variance estimates, we use the variance error estimates derived in the first step (4). We decide to refer to the RPRI (3), even if is not widely used, because it can conceptually be an element of connection between the first and second stage

of Two-step procedure. In fact it is calculated together with the measures in the first step and it determines the magnitude of measurement error in the second one.

In **RSM-LRM**, for the second step we refer to a simple linear regression model, where the measures are used directly in the regression equation, without considering that they are affected by measurement errors:

$$Y_1^* = \beta_2 Y_2^* + u_1.$$

What we expect, and what we will verify with the simulation study, is that estimated regression coefficient, obtained with the latter method, is lower than that obtained with the **RSM-LRMme** method, because of the attenuation effect of measurement error (Fuller, 1987).

4. The case study

4.1. The simulation design

The objective of this study is to compare the results obtained applying, in the same situations, the 4 different analytical methods, previously described. We created many different simulated datasets in order to evaluate the obtained estimates, knowing the real value of the parameter of interest.

We consider m latent variables and p indicators, with p_j that refers to the number of indicators of the j -th latent variable.

We considered two scenarios: the first is simple with only two latent factors and a dependence relationship of the first versus the second one; in the second scenario we considered three latent factors, where the first is dependent from the other two. Each indicator is a categorical variable with 5 ordered categories. The data generating model is (5-7).

For each scenario, we fixed the variance of the latent variables equal to 1, then we changed the value of three basic model parameters in order to create different configuration sets.

1. We changed p_2 , the number of indicators for the first independent regressor. From the IRT (Baker and Kim, 2004), we know that increasing the number of indicators, the measure reliability increases; we want to control if it is verified in our simulations.
2. Fixing at 1 the variance of the dependent latent variable θ_1 , we change its structural error variance, ψ_1 . In this way, we can define the strength of the dependence link between the latent variables: increasing ψ_1 , we reduce the dependence relationship between of θ_1 from other two independent latent variables.
3. Only for the second scenario, we changed the intensity of dependence of θ_1 from the other two independent latent variables θ_2 and θ_3 . We have

$$\theta_1 = \beta_2 \theta_2 + \beta_3 \theta_3 + \zeta_1.$$

Whereas the structural errors (ζ_h) are uncorrelated with each other, we know that

$$\theta_1 = \beta_2 \theta_2 + \beta_3 \theta_3 + \zeta_1.$$

Because of $\text{var}(\theta_1) \equiv 1$, $\kappa = (1 - \psi_1)$ represents the dependence of θ_1 from the regressors.

Obviously, when we change these three groups of parameters, we change also the values of the coefficients β_2 and β_3 . The following table summarizes the values we assigned to all parameters, arranged in different configurations for the analysis.

p_i	$p_1 = 5$ $p_2 = 5, 10$ $p_3 = 5$ (only for the second scenario)
ψ_1	0.1, 0.3, 0.5, 0.7, 0.9
β	First scenario $\beta_2 = 0.95, 0.84, 0.71, 0.55, 0.32$ Second scenario $\beta_2, \beta_3 = 0.90, 0.79, 0.67, 0.59, 0.52, 0.50, 0.39, 0.30, 0.22, 0.17, 0.10$
dependence	For the second scenario, we create 3 different situations: <i>dep1</i> : $\beta_2 = \sqrt{\frac{\kappa}{10}}$ and $\beta_3 = 3\beta_2$. <i>dep2</i> : $\beta_2 = \beta_3 = \sqrt{\frac{\kappa}{2}}$. <i>dep3</i> : $\beta_2 = 3\sqrt{\frac{\kappa}{10}}$ and $\beta_3 = \sqrt{\frac{\kappa}{10}}$ (only when $p_2 = 10$).

In total, we created 35 different parameter configurations, 10 for the first scenario and 25 for the second one. For each set, we simulate N samples of size n . We make some preliminary studies to understand the optimal number of samples (N) and the sample size (n). Because of in the SEM context large samples methods are used, we fixed the sample size to 1000. To decide the number of samples, we tested the stability of estimates for each of the 4 different methods described in section 3 and we noted that estimates become stable with a few dozen repetitions. Focusing on the standard error of the estimates more iterations are needed, so we set the number of samples N equal to 500, for each parameter set.

The data have been generate with *Mplus* (Muthén and Muthén, 2007), considering all the configurations described in the previous section. We started generating multivariate normal data for the independent variables in the model $(\theta_2, \theta_3, \Delta_1)$. Then the data for the continuous dependent variable, θ_1 , have been generated according to a distribution that is multivariate normal conditional on the independent variables. Finally, we generated the categorical dependent variables, Y_j , according to the probit model, using the fixed values of the thresholds and item difficulty parameters. The thresholds and the item difficulties, δ_j , have been chosen to obtain items with different (symmetric and asymmetric) response distributions (see Figure 1).

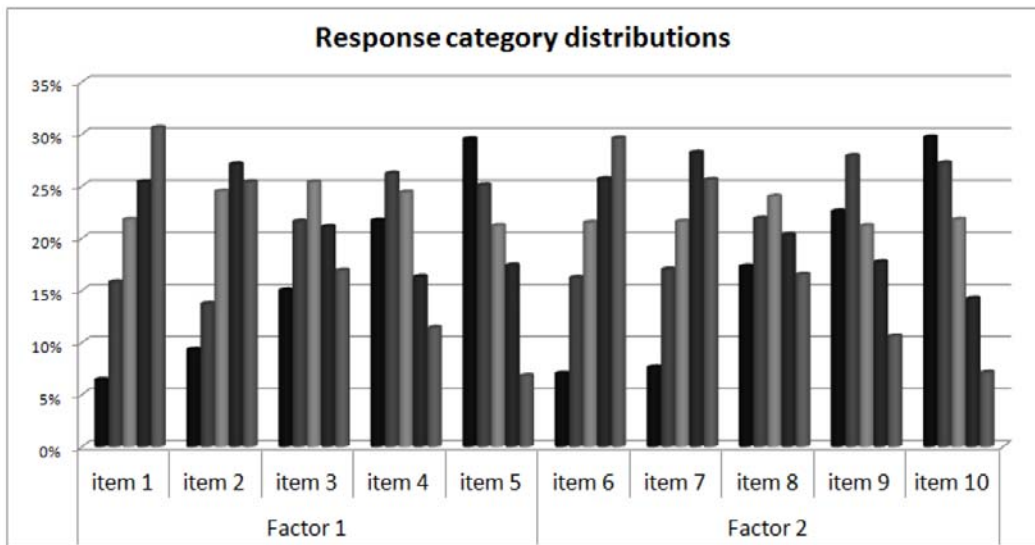


Figure 1. The frequency distributions for the 5 items response categories of the 10 items of the first scenario, $p_2 = 5, \psi_1 = 0.1$.

4.2. The simulation results

In this Section, we will present the key findings obtained from our simulation study. All reported results refer to b_2 and b_3 , the estimates of the β coefficients that indicate the dependence relationships between the latent variables, the final goal of many of the socio-economic studies. To compare the results obtained with the 4 estimation methods, we evaluate:

- the Relative Bias, $RB(b) = \frac{\beta - b}{\beta}$;
- the Relative Standard Error, $RSE(b) = \frac{SE(b)}{\beta}$;
- the Relative Root Mean Square Error, $RMSE(b) = \sqrt{RB(b)^2 + RSE(b)^2}$.

Because of the β coefficients takes values in [0.1, 0.9], we divided the indices by the actual value of the parameter, to allow a correct comparison in the several presented cases.

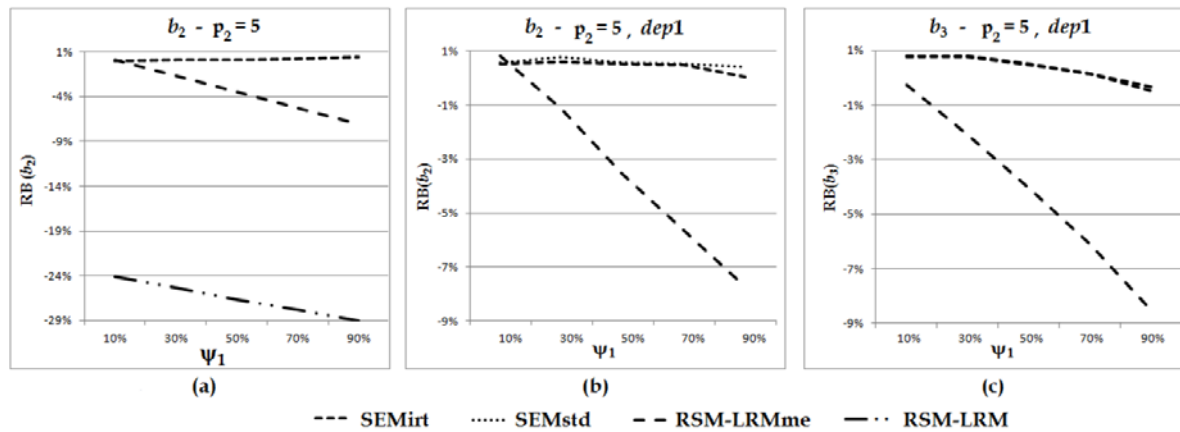


Figure 2. Simulation results for the Relative Bias (RB) in percentage of β coefficient estimates obtained from the One-step and the Two-step procedures.

In Figure 2, we see the relative bias for the β coefficients. In graph (a), we present the results for all 4 methods in first scenario, $p_2 = 5$. We can immediately observe that the results of **RSM-LRM** show a strong negative bias, consistent with the theory of measurement errors; for this reason, we do not consider it in the following results. In graphs (b) e (c), we brought the results for the second scenario, case $p_2 = 5$ dep1, for b_2 and b_3 , respectively. The two One-step procedures, that follow a similar trend, show a distortion of reduced entity (in absolute value less than 1%). The bias of the **RSM-LRMme** increases as ψ_1 increases, but is always lower than 10%, in the extreme case too.

Looking at the relative standard error, we have seen that the **RSM-LRM** method shows $RSE(b)$ lower than other procedures, even if they are all very close. For all the 4 methods, the standard errors increase as ψ_1 increases.

To assess the impact of bias and error standard together, we refer to the Relative Root Mean Square Error. In Table 1 we report the $RMSE(b)$ for the estimates of β coefficients. For synthesis, we does not report the data referring to the **RSM-LRM**. It presents a very high RMSE (up to 3 times that of other methods), due to strong bias already seen in previous graphs, not sufficiently compensated by the good accuracy of the estimation.

Table 1 Simulated RMSE in percentage of β coefficient estimates obtained from the **SEMstd** and the **RSM-LRMme** estimation methods for all the different parameter configurations

	Case	Parameter	Procedure	Ψ_1					
				0.10	0.30	0.50	0.70	0.90	
First scenario	$p_2 = 5$		SEMstd	1.50	2.30	3.52	5.45	11.02	
			RSM-LRMme	1.88	3.23	5.39	8.03	13.93	
	$p_2 = 10$		SEMstd	1.17	2.06	3.34	5.43	11.14	
			RSM-LRMme	1.53	3.05	4.95	7.63	13.56	
Second scenario	$p_2 = 5$	dep1	b ₂	SEMstd	8.53	10.87	14.24	20.12	37.09
			RSM-LRMme	9.69	12.16	16.26	23.07	41.32	
		b ₃	SEMstd	1.83	2.74	3.86	5.91	11.92	
			RSM-LRMme	2.14	3.71	6.01	9.08	15.91	
	dep2	b ₂	SEMstd	3.44	4.58	5.98	8.59	16.56	
			RSM-LRMme	3.85	5.24	7.65	11.27	19.95	
		b ₃	SEMstd	3.34	4.46	5.83	8.44	16.37	
			RSM-LRMme	3.71	5.42	7.80	11.49	20.30	
	$p_2 = 10$	dep1	b ₂	SEMstd	8.24	10.42	13.57	18.93	35.30
				RSM-LRMme	9.65	11.96	15.43	21.76	39.03
			b ₃	SEMstd	1.84	2.78	3.66	5.74	11.70
				RSM-LRMme	2.18	3.70	5.84	8.88	15.07
		dep2	b ₂	SEMstd	3.02	4.11	5.48	8.09	15.67
				RSM-LRMme	3.39	4.57	6.49	9.75	17.97
			b ₃	SEMstd	3.28	4.30	5.63	8.21	16.18
				RSM-LRMme	3.73	5.39	7.70	10.98	19.37
	dep3	b ₂	SEMstd	1.41	2.42	3.53	5.53	11.34	
			RSM-LRMme	1.69	3.10	4.93	7.64	13.78	
		b ₃	SEMstd	7.62	10.33	13.65	19.39	37.36	
			RSM-LRMme	8.72	11.68	15.79	21.96	40.88	

With regard to the One-step procedures, **SEMirt** and **SEMstd**, the results are essentially identical, differing only at 3 or 4 decimal places, so we report only the results of more general method **SEMstd**. Consistent with the assumptions, the One-step procedures have the lowest **RMSE(β)**, consequence of the simultaneous estimation of all parameters of the model.

The method **SEM-LRMme** has low relative RMSE, very close to the **SEMstd**. This result is crucial for our analysis. Looking at the data, in fact, we note that the proposed Two-step method does not introduce a strong source of error in the model, even if it divides the estimated parameters in two distinct phases; the results obtained are indeed very close to those of the One-step procedures (only from 2 to 4 percentage points more). RSME can assume high values when the variance of the structural error of θ_1 is high (at least 0.7) and θ_1 has a strong bond of dependency with θ_2 or θ_3 , case **dep1** or **dep3** respectively.

One last thing to consider is the reliability of the obtained measures. In both scenarios, RPRI and α are perfectly consistent and all the index values are significant because they are greater than 0.7 (Nunnally and Bernstein, 1994). As mentioned above, α values are always greater than RPRI; for computing the value of α , we can just multiply the

RPRI index by a factor g (slightly different for each combination of parameters). It's also interesting to see that, consistent with the literature, the measures built across 10 indicators are characterized by an index of reliability far greater than the others.

4.3. The real data results

In this section we apply the described methods to a real case: a study addressing the quality of work of a sample of employees in the Italian social cooperatives, named *ICSI*²⁰⁰⁷ (Carpita, 2009; Carpita and Golia, 2011). The data have been collected through a questionnaire designed to investigate different constructs, including: job satisfaction, motivation, job complexity (perceived activities), procedural fairness (existence of transparent of rules that governs the relationship between worker and cooperative), organizational fairness (perception of the worker in relation to their working conditions and its participation in organizational life) and distributive fairness (distributing resources, balance between what the worker gives the organization and what that it receives).

Besides getting a good measure of these constructs, we are obviously interested in understanding the relationships between them. We have focused our attention on three latent constructs: distribution fairness (*DF*), procedural fairness (*PF*) and overall satisfaction (*OS*). Referring to the preliminary analysis carried out by Carpita and Golia (2011), we used 7 items for *DF*, 8 items for *PF* and 11 items for *OS*. We want to understand the relationship between the overall satisfaction and the other two constructs, so that the structural model is:

$$OS = \beta_2 DF + \beta_3 PF + \zeta_1.$$

Table 2. Real data analysis: β coefficient and standard error estimates obtained from the 4 described estimation methods.

Procedure	b_2	$SE(b_2)$	b_3	$SE(b_2)$	$\hat{\psi}_1$
SEMstd	0.054	0.020	0.821	0.013	0.33
SEMirt	0.068	0.020	0.795	0.013	0.36
RSM-LRMme	0.084	0.013	0.725	0.013	0.46
RSM-LRM	0.077	0.013	0.661	0.013	0.56

In Table 2 we report the results obtained with the 4 discussed procedures. All 4 procedures provide similar information about the intensity of the relationship between the 3 latent constructs. In particular the 4 methods showed a positive but weak effect of *DF* on *OS* (coefficient estimates are between 0.05 and 0.08), while there is a strong and positive effect of *PF* on *OS* (coefficient estimates are between 0.66 and 0.82). The standard errors of the estimates are very low and roughly the same. Focusing on this second regression coefficient, we see that the Two-step procedure estimates are lower than those of One-step procedure estimates. **RSM-LRM** is strongly unbiased (the value is the lowest); **RSM-LRMme** reduces the attenuation due to error of measurement, but it still has a certain level of bias ($RE(b)\% = 10.98\%$, considering the **SEMstd** estimate as the closest to the actual value of the parameter).

The RPRI, used in the Two-step procedure, is equal to 0.9 for the *DF* measure, 0.86 for the *PF* measure and 0.89 for the *OS* measure.

5. Conclusions

A first consideration is about the **RSM-LRM** method. Although sometimes the standard linear regression is used also with variables affected by measurement error, our simulation showed that the estimator bias for the parameters of interest is very strong.

Remembering that one of our purpose was to implement a Two-step procedure efficient and precise, we focuses the attention on the Root Mean Square Error index, that combines an assessment of bias and efficiency. From the simulation results, we have that the Two-step procedure has a slight distortion and a loss of efficiency, but its estimates are coherent with those provided by the One-step procedure. It is a very interesting result, that provides a useful tool for future analysis with real data.

We mentioned the advantages of IRM in terms of greater flexibility of analysis and possibility to check the hypothesized relations, but we did not know what was the price to pay in terms of loss of efficiency and distortion. Given this simulation data, we could say that, for the cases presented, the Two-step procedure results sufficiently precise and unbiased.

Obviously the choice of which procedure to implement is prerogative of the researcher and it depends strongly of the analysis purposes, but for the cases described in the simulation study, both the two approaches could be used to obtain statistically useful results.

Moreover, the simulation allowed us to compare the performance of two reliability indices: Cronbach's alpha and Rasch Person Reliability Index. The results showed that these indices followed a similar pattern, thus providing similar indications. RPRI is more precautionary because it is always lower than α : so RPRI assigns a greater value to measurement error. In our **RSM-LRMme** method, we decide to use the RPRI as it is the natural index of reliability in the IRT (Schumacker and Smith, 2007). In fact, this index represents the logical link between the first and second steps of our estimation procedure. In the first step we get, through a RSM, the estimates of Rasch measures and, through RPRI, their reliability. In the second step, we develop a regression model with these measures, using a function of RPRI as estimate of the variance of their errors. It remains an open question whether and how we can check analytically the measure reliability in SEM.

References

1. Baker, F. B. and Kim, S. H. **Item Response Theory: Parameter Estimation Techniques**. CRC Press, New York, 2nd edition, 2004
2. Cagnone, S., Mignani, S. and Moustaki, I. **Latent variable models for ordinal data**, in Monari, P., Bini, M., Piccolo, D., and Salmaso, L. (eds.) "Statistical Methods for the Evaluation of Educational Services and Quality of Products, Contributions to Statistics", Physica-Verlag HD, 2010, pp. 17-28
3. Carpita, M. and Golia, S. **Measuring the quality of work: the case of the Italian social cooperatives**, Quality & Quantity, Springer Netherlands, 2011, pp. 1-27
4. Carpita, M. **La qualità del lavoro nelle cooperative sociali. Misure e modelli statistici**, FrancoAngeli, Milano, 2009
5. Cronbach, L. **Coefficient alpha and the internal structure of tests**, Psychometrika, 16(3), 1951, pp. 297-334
6. De Vellis, R. **Scale Development: Theory and Application**, Sage Publications, Newbury Park, 1991

7. Fox, J. **Teacher's corner: Structural equation modeling with the sem package**, R. Structural Equation Modeling: A Multidisciplinary Journal, 13(3), 2006, pp. 465-486
8. Fuller, W. **Measurement error models**, John Wiley & Sons, New York, 1987
9. Furr, R. M. and Bacharach, V. R. **Psychometrics: An Introduction**, Thousand Oaks, CA: Sage, 2008
10. Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., and Stover, A. **Full-information item bifactor analysis of graded response data**, Applied Psychological Measurement, 31(1), 2007, pp. 4-19
11. Hays, R.D., Morales, L.S. and Reise, S.P. **Item response theory and health outcomes measurement in the 21st century**, Medical Care, 38(9 Suppl): II, 2000, pp. 28-42
12. Linacre, J. M. **Kr-20 / Cronbach alpha or Rasch reliability: Which tells the truth?**, Rasch Measurement Transactions, 11(3), 1997, pp. 580-581
13. Mair, P. and Hatzinger, R.. **Extended Rasch modeling: The erm package for the application of IRT models**, R. Journal of Statistical Software, 20(9), 2007, pp. 1-20
14. Muthén, B. **A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators**, Psychometrika, 49(1), 1984, pp. 115-132
15. Muthén, L. K. and Muthén, B. O. **Mplus statistical analysis with latent variables. Users Guide**, Muthén & Muthén, Los Angeles, 2007
16. Nunnally, J. C. and Bernstein, I. H. **Psychometric Theory**, McGraw-Hill, New York, 3rd edition., 1994
17. Rasch, G. **On general laws and the meaning of measurement in psychology**, in "Proceedings of the IV. Berkeley Symposium on Mathematical Statistics and Probability", University of California Press, Berkeley, Vol. IV, 1961, pp. 321-333
18. Schumacker, R. E. and Smith Jr., E. V. **Reliability - a Rasch perspective**, Educational and Psychological Measurement, 67(3), 2007, pp. 394-409

¹ In 2006 Anna Simonetto holds a bachelor's degree in "Applied Science to Economics" at the University of Brescia, with a thesis entitled: "*Data mining models for predicting customer behavior: the case Bresciaonline*". In 2010 she earned her Ph.D. in "Statistics and Applications" at the University of Milano-Bicocca with a thesis on: "*Estimation Procedures for Latent Variable Models with Psychological Traits*".

She is currently a research fellow at the University of Brescia. Her research interests encompass: multivariate analysis of large databases, algorithmic techniques for data analysis, latent variable models.