UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA

Scuola di Dottorato in Statistica e Matematica applicata alla Finanza

Dottorato di Ricerca
in
Statistica ed Applicazioni

XXII ciclo

Estimation procedures for latent variable models with psychological traits

Relatore:

Chiar.mo Prof. Maurizio Carpita

Dottoranda:

Anna Simonetto

A.A. 2008-09

If you have an apple and I have an apple,
and we exchange apples,
we both still only have one apple.
But if you have an idea and I have an idea,
and we exchange ideas,
we each now have two ideas.
(George Bernard Shaw)

Contents

Introduction

The starting point for this thesis is a concrete problem: to measure, using statistical models, aspects of subjective perceptions and assessments and to understand their dependencies. The objective of the research is to study the statistical properties of some estimators of the parameters of regression models with variables affected by measurement errors. These models are widely used in surveys based on questionnaires developed to detect subjective assessments and perceptions with Likert-type scales. It is a highly debated topic, as many of the relevant aspects in this field are not directly observable and therefore the variables used to estimate them are affected by measurement errors. The models with measurement errors were very thorough in literature. Already addressed in the middle of last century by Fuller (1987), this theme has been widely echoed by many researcher that investigated different aspects of this problem, with respect to their specific applications.

In literature we can identify several areas of research, based on the different approaches that the authors have with this topic; in this work we will developed two of the most used. Obviously, according to the approach chosen, different models were proposed to estimate the relationships between variables affected by measurement error. After exposing the main features of these models, the thesis focuses on providing an original contribution to comparative analysis of the two presented approaches.

We started from a preliminary critical reading of the methodological literature on three specific multivariate models used in the context outlined above: the Measurement Error Model (MEM), the Item Response Model (IRM) and the Structural Equation Model (SEM).

In the first Chapter, we will present a review of major works concerning

the general theme of MEMs, inspired by one of the basic texts (Fuller, 1987), then integrating with more recent theoretical contributions (Carroll et al., 2006). We will present only some basic regression models, mainly linear with a little hint at the nonlinear case, and two methods of estimation (Regression-calibration and SIMEX).

In the second Chapter, we will deepen two more modern approaches to the problem of measurement error in social sciences, the IRM and the SEM. After a brief overview of the literature, we will present some famous models for each approach, to focus, in the last sections of the chapter, on two different estimation procedures. The first procedure, named One-step, considers simultaneously all the parameters involved in the complete model for the hypothesized latent and observed variables (using the SEM). The second procedure, named Two-step, starts obtaining, by the observed variables, the measures associated to latent variables (using the IRM). In a second phase, we derive parameter estimates of the regression model assumed for the latent variables, using in its specification the measures of the first step and considering that they are affected by measurement errors (using the MEM).

The properties of the Two-step procedure have not yet been adequately detailed in the literature and we did not find significant contributions in this regard. The simulation study will indicate a possible analysis strategy, that could provide some useful prescriptive about it. One of the most innovative aspects of the proposed procedure concerns the method of the measurement error inclusion in the second step. Some authors, after obtaining the measures in the first step, estimate the regression model without taking into account that the model variables are affected by measurement error.

The simulation study want 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 also evaluating that it allows better control in both phases: measures construction and regression interpretation.

Another original aspect of the thesis concerns the reliability index used to estimate the variance of measurement error: the Rasch Person Reliability

Index. This index is still not widely used in literature, but it can conceptually be an element of connection between the first and second stage of Two-step procedure, as it is calculated together with the measures in the first step and it determines the magnitude of measurement error in the second one.

In the final Chapter, we will present the simulation study, where we compare the results obtained through an articulated process of analysis. This original comparison, conducted with reference to the usual properties of estimators (bias, standard error and mean square error) is of particular interest especially from the practical point of view. In fact, we know that, compared to the One-step, the Two-step procedure allows greater flexibility of analysis and possibility of verification of hypothesized relations, especially for more complex models consisting of several latent variables and observed variables, but it is less precise and it is, in some extend, biased. The simulation study want to investigate this bias and loss of precision.

Chapter 1

Measurement Error in Regression Models

1.1 Introduction

In this chapter, we will present the theory of regression models for variables affected by measurement errors. This may arise for a number of reasons, for example it could be due to bad measurement tools or just because the true variable can not be measured directly. When we have discrete variables, the measurement error is indicated as misclassification. In the first Section, we will make a general presentation of the topic, introducing some useful definitions to understand the models presented later. In Section 1.3, we will present the measurement error linear regression model (the univariate case and, briefly, the extension to multivariate case), while the response variable error model is shown in Section 1.4. Finally, we present two different estimation methods, the Regression-Calibration (Section 1.5) and the Simulation-Extrapolation (Section 1.6).

1.2 A General Overview of Measurement Error Models (MEMs)

We will consider fitting models that relate q ($q \geq 1$) response variables $\mathbf{Y} = (Y_1, \ldots, Y_q)$, to m ($m \geq 0$) error-free predictors $\mathbf{Z} = (Z_1, \ldots, Z_m)$ and p ($p \geq 1$) true predictors $\mathbf{X} = (X_1, \ldots, X_p)$. We can not observe directly the true predictors (called *latent variables*), but instead we have the observed variables $\mathbf{W} = (W_1, \ldots, W_p)$. For this reason, often we are not able to estimate parameters of interest given only the information contained in the sample of (**Y, W, Z**) values. In fact we want to fit statistical models to data formulated in terms of well-defined but unobservable $\mathbf{X}$, using information on measurements $\mathbf{W}$ that are only correlated with $\mathbf{X}$.

First of all we have to distinguish between two types of error:

- the *classical additive error*

$$\mathbf{W} = \mathbf{X} + \mathbf{U} \qquad E(\mathbf{U}|\mathbf{X}) = 0, \tag{1.1}$$

  where the observed variable $\mathbf{W}$ is equal to the true one plus a measurement error and so its variability is greater than that of $\mathbf{X}$;

- the *Berkson error*

$$\mathbf{X} = \mathbf{W} + \mathbf{U} \qquad E(\mathbf{U}|\mathbf{W}) = 0, \tag{1.2}$$

  where the true variable $\mathbf{X}$ varies around $\mathbf{W}$ and so its variability is greater than that of the observed variable.

When we start an analysis with measurement error models, it is important to understand which of these two types of error represents better the real situation, because "*for a given measurement error variance, if you want to convince yourself that you have lots of statistical power despite measurement error, just pretend that the measurement error is Berkson and not classical*" (Carroll et al, 2006). If one assumes, for example, that the measurement error is the classical type instead of Berkson type, then the variance of $\mathbf{X}$ is, from

(1.2), the variance of $\mathbf{W}$ minus the variance of the classical measurement error $\mathbf{U}$, while if you assume a Berkson error model the variance $\mathbf{X}$ is much greater: var[$\mathbf{X}$]=var[$\mathbf{W}$]+var[$\mathbf{U}$]. In practical analysis, the choice is quite simple: we have a Berkson error when for all individuals in a group we have the same value of the error-prone covariate, but the true value is particular to each individual, for example miners working in the same sector for the same period are assigned the same exposure to dust, even if the true exposure is obviously different for each of them. We choose the classical error model if an error-prone variable is necessarily measured uniquely to a subject, and that measurement can be replicated, a typical example is the blood pressure measurement.

Focusing the attention on the true variable $\mathbf{X}$, in literature we find two different approaches: *classical functional models*, where $\mathbf{X}$ is regarded as a sequence of unknown fixed constants or parameters, and *classical structural models*, where $\mathbf{X}$ is considered as random variables. In this work we will use the distinction, used by Carroll et al. (2006), between:

- the *functional modeling*, if we make only minimal assumptions about the distribution of $\mathbf{X}$ (that could be either fixed or random);
- the *structural modeling*, where we place parametric models on the distribution of the random $\mathbf{X}$.

If we focus on the specification of the model, we can make a similar distinction between:

1. *Error models*, where we model the conditional distribution of $\mathbf{W}$ given $(\mathbf{Z}, \mathbf{X})$;
2. *Regression calibration model*, where we model the conditional distribution of $\mathbf{X}$ given $(\mathbf{Z}, \mathbf{W})$.

In the first group, that includes the classical measurement error model (1.1), we will suppose that the relationship between $\mathbf{W}$ and the unobserved $\mathbf{X}$ also depends on the other predictors $\mathbf{Z}$:

$$\mathbf{W} = \gamma_x' \mathbf{X} + \gamma_z' \mathbf{Z} + \mathbf{U}, \qquad E(\mathbf{U}|\mathbf{X}, \mathbf{Z}) = 0.$$

In the second group, we want to model the distribution of the unobserved explanatory variables as a function of the observed ones:

$$\mathbf{X} = \delta'_w \mathbf{W} + \delta'_z \mathbf{Z} + \mathbf{U}, \qquad E(\mathbf{U}|\mathbf{W}, \mathbf{Z}) = 0.$$

If we assume $\delta_w = 0$ and $\delta_z = 1$, we can see that this category includes the Berkson error model (1.2).

Analyzing the conditional density used in these models, it is possible to find an interesting relationship between error models and regression calibration models. It permits to find a model for the distribution of $\mathbf{X}$ given $\mathbf{W}$, starting with a model for the distribution of $\mathbf{W}$ given $\mathbf{X}$. If we have a Structural Model, and so we know the marginal distribution of $\mathbf{X}$, then we can convert an error model in a regression calibration model using the Bayes theorem:

$$f_{\mathbf{X}|\mathbf{W}}(x|w) = \frac{f_{\mathbf{W}|\mathbf{X}}(w|x) f_{\mathbf{X}}(x)}{\int f_{\mathbf{W}|\mathbf{X}}(w|x) f_{\mathbf{X}}(x) dx} ,$$

where $f_{\mathbf{X}}$ is the density function of $\mathbf{X}$, $f_{\mathbf{W}|\mathbf{X}}$ is the density function of $\mathbf{W}$ given $\mathbf{X}$ and $f_{\mathbf{X}|\mathbf{W}}$ is the density function of $\mathbf{X}$ given $\mathbf{W}$.

Another important distinction is between *differential* and *nondifferential* measurement error. We have nondifferential measurement error if $\mathbf{W}$, conditionally to the information contained in $\mathbf{X}$ and $\mathbf{Z}$, doesn't bring additional information to $\mathbf{Y}$. A nondifferential measure contains no more information for the prediction of $\mathbf{Y}$ than those already contained in $\mathbf{Z}$ and $\mathbf{X}$. If $f_{\mathbf{Y}|\mathbf{ZXW}} = f_{\mathbf{Y}|\mathbf{ZX}}$, we call $\mathbf{W}$ a surrogate for $\mathbf{X}$, and in this case, even if $\mathbf{X}$ is not observable, it is possible to estimate the model parameters of response given the true covariate.

With differential measurement error, generally we can not do that because $\mathbf{Y}$ is not conditionally independent of $\mathbf{W}$ given the true covariates. In this case, for example, we need a validation subsample. in which both the measured value and the true value are recorded. We find this particular data structure when, for example, $\mathbf{W}$ is not only a mismeasured version of $\mathbf{X}$, but it is a separate variable acting as a proxy of $\mathbf{X}$. Often the surrogate

status of $\mathbf{W}$ depend on the other variables in the model. For example, let $X \sim N(\mu_x, \sigma_x^2)$, we assume that $\epsilon_1, \epsilon_2, U_1, U_2$ are normal random variables with zero mean, mutually independent and all independent from X. If

$$\begin{aligned} Z &= X + \epsilon_1 + U_1, \\ Y &= b_x + b_z Z + b_x X + \epsilon_2, \\ W &= X + \epsilon_1 + U_2, \end{aligned}$$

then

$$\mathrm{E}[Y|X] = \mathrm{E}[Y|X, W],$$

but

$$\mathrm{E}[Y|Z, X, W] \neq \mathrm{E}[Y|Z, X],$$

because the measurement error $(W - X)$ is correlated with the covariate Z.

We can better understand the importance of nondifferential measurement error observing the relationship between $\mathbf{Y}$ and $\mathbf{W}$ in simple linear regression:

$$\begin{aligned} E[\mathbf{Y}|\mathbf{W}] &= E\{E[\mathbf{Y}|\mathbf{X}, \mathbf{W}]|\mathbf{W}\} && \text{(1.3a)} \\ &= E[E(\mathbf{Y}|\mathbf{X})|\mathbf{W}] && \text{(1.3b)} \\ &= E[b_0 + b_x\mathbf{X}|\mathbf{W}] && \text{(1.3c)} \\ &= b_0 + b_x E[\mathbf{X}|\mathbf{W}]. && \text{(1.3d)} \end{aligned}$$

Looking at the previous equations, we can interpret the regression in the observed data as a linear regression of $\mathbf{Y}$ on $E(\mathbf{X}|\mathbf{W})$, but this is true if and only if we have nondifferential measurement error, otherwise the passage (1.3b) is not verified.

1.3 The Linear Regression Model (LRM)

1.3.1 A Single Explanatory Variable

In this section we will briefly present the effects of measurement error in a normal simple linear regression model. We begin from the classical linear

regression model with one independent variable:

$$Y = b_0 + b_x X + \epsilon, \tag{1.4}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$.

The *naive* ordinary least squares estimator, that is the usual method for error-free data, is

$$\hat{b}_x = \left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{-1} \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}),$$

and it is the best linear unbiased estimator (BLUE) for b_x (Faliva, 1987).

For the following analysis, we consider the classical additive measurement model (1.1) with

$$\begin{pmatrix} X \\ \epsilon \\ U \end{pmatrix} \sim \text{NI}\left[\begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}\right].$$

If we assume that Y is defined by (1.4), than the vector (Y, X) follows a normal distribution with mean

$$E\begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix} = \begin{bmatrix} b_0 + b_x\mu_x \\ \mu_x \end{bmatrix} \tag{1.5}$$

and covariance matrix

$$\begin{bmatrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{bmatrix} = \begin{bmatrix} b_x^2\sigma_x^2 + \sigma_\epsilon^2 & b_x\sigma_x^2 \\ b_x\sigma_x^2 & \sigma_x^2 \end{bmatrix}.$$

Moreover, when we include W, that is defined by (1.1) and it is jointly normally distributed with (Y, X), we obtain a multivariate normal model for

(Y, X, W):

$$\begin{pmatrix} Y \\ X \\ W \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} b_0 + b_x\mu_x \\ \mu_x \\ \mu_w \end{pmatrix}, \begin{pmatrix} b_x^2\sigma_x^2 + \sigma_\epsilon^2 & b_x\sigma_x^2 & b_x\sigma_{xw} + \sigma_{\epsilon w} \\ b_x\sigma_x^2 & \sigma_x^2 & \sigma_{xw} \\ b_x\sigma_{xw} + \sigma_{\epsilon w} & \sigma_{xw} & \sigma_w^2 \end{pmatrix} \right].$$

As we are in the context of measurement error models, the only available data are the observed values of (Y, W), so the relevant sampling model is the marginal distribution of (Y, W):

$$\begin{pmatrix} Y \\ W \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} b_0 + b_x\mu_x \\ \mu_w \end{pmatrix}, \begin{pmatrix} b_x^2\sigma_x^2 + \sigma_\epsilon^2 & b_x\sigma_x^2 \\ b_x\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix} \right],$$

where $\mu_w = \mu_x$.

With this error structure, the regression coefficient becomes

$$\hat{b}_{x*} = \left[\sum_{i=1}^{n} (w_i - \bar{w})^2 \right]^{-1} \sum_{i=1}^{n} (w_i - \bar{w})(y_i - \bar{y}) = \frac{\sigma_{wy}}{\sigma_w^2}, \tag{1.6}$$

from which, by the properties of the bivariate normal distribution, we obtain that the measurement error attenuates the regression coefficient

$$E\{\hat{b}_{x*}\} = b_x \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} = b_x \frac{\sigma_x^2}{\sigma_w^2} = b_x\lambda.$$

It's a well-known achievement by which we can demonstrate that the least squares regression coefficient for the bivariate model with independent measurement error in W is biased toward zero and we call

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < 1, \tag{1.7}$$

as *reliability ratio*.

Often we can assume to know σ_ϵ^2, for example because we made a large number of independent repeated measurements, and we define the sample estimators of $(\sigma_y^2, \sigma_{wy}, \sigma_w^2)$ as $(\hat{\sigma}_y^2, \hat{\sigma}_{wy}, \hat{\sigma}_w^2)$.

The population squared correlation between X and Y is

$$R^2_{xy} = \frac{\sigma^2_{xy}}{\sigma^2_x \sigma^2_y} = b_x \frac{\sigma^2_x}{\sigma^2_y} \ , \tag{1.8}$$

and the population squared correlation between W and Y is

$$R^2_{wy} = \frac{\sigma^2_{wy}}{\sigma^2_w \sigma^2_y} = \lambda R^2_{xy} \ . \tag{1.9}$$

Comparing (1.8) and (1.9), we can see that the introduction of independent measurement error induces a reduction of the squared correlation proportional to the reliability ratio λ.

1.3.2 Bias Correction

If we know the reliability ratio (1.7), it is possible to construct an unbiased estimator of structural regression coefficient b_x, that relates Y to the true value of X in the model (1.4):

$$\hat{b}_{x**} = \frac{\hat{b}_{x**}}{\lambda} \ , \tag{1.10}$$

with $E[\hat{b}_{x**}] = b_x$. Knowing λ, an estimator of the squared correlation between X e Y is

$$\hat{R}^2_{xy} = \frac{\hat{R}^2_{wy}}{\lambda} \ ,$$

where

$$\hat{R}^2_{wy} = \frac{\hat{\sigma}^2_{wy}}{\hat{\sigma}^2_w \hat{\sigma}^2_y} \ .$$

In practice, it is very difficult to known the reliability ratio and so we have to estimate it with

$$\hat{\lambda} = \frac{\hat{\sigma}^2_w - \hat{\sigma}^2_u}{\hat{\sigma}^2_w} \ .$$

The problem is that when we reduce the bias, the variance increases: the method-of-moments estimator (1.10) has variance $Var[\hat{b}_{x**}] = \sigma^2_*/\lambda^2$, where σ^2_* is the variance of (1.6), that is greater than the variance of the

biased estimator $\hat{b}_{x*}$, because $\lambda < 1$. Analyzing the residual variance of the regression of Y on W

$$\text{var}[Y|W] = \sigma_\epsilon^2 + b_x^2\sigma_u^2\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2} = \sigma_\epsilon^2 + \lambda b_x^2\sigma_u^2,$$

we can better understand that the measurement error causes an increase of variability in the data. If we substitute $(X = W - U)$ into the regression model, we obtain a new model

$$Y = b_0 + b_x W + (\epsilon - b_x U),$$

with covariate W and error $(\epsilon - b_x U)$ that has variance $(\sigma_\epsilon^2 + b_x^2\sigma_u^2) > \sigma_\epsilon^2$. The bias is due to the fact that the error and the covariate have a common component U, so they are not uncorrelated.

To better understand the *bias versus variance tradeoff* we can use an index that considered simultaneously the variance and the bias, as the means squared error (MSE), that is the sum of variance plus the squared of bias:

$$\begin{aligned} \text{MSE}(\hat{b}_{x*}) &= \sigma_*^2 + (1-\lambda)^2 b_x^2 \\ \text{MSE}(\hat{b}_{x**}) &= \frac{\sigma_*^2}{\lambda^2}\ . \end{aligned}$$

From the above equations, we can deduce that

$$\text{MSE}(\hat{b}_{x**}) < \text{MSE}(\hat{b}_{x*}) \quad \text{iff} \quad \sigma_*^2 < \frac{\lambda^2(1-\lambda)b_x^2}{1+\lambda}\ .$$

Because σ_*^2 decreases with increasing sample size, we can say that, in large samples, it is opportune to correct for attenuation due to measurement error.

1.3.3 The Simple LRM with Different Error Structure

Now we consider the impact of different measurement error structures on the simple regression model (1.4). We start from the Berkson error $(X = W + U)$,

where we have

$$\begin{aligned} E(X|W) &= W \\ E(Y|W) &= b_0 + b_x W \ , \\ \sigma_{xw} &= \sigma_w^2 \\ \sigma_x^2 &= \sigma_w^2 + \sigma_u^2. \end{aligned}$$

In this case the naive estimator that regress Y on W is unbiased for b_0 and b_x, but we have an increase in the residual variance and a corresponding decrease in the model coefficient of determination (R^2).

When we have a model with differential error, W is not unbiased for X and we can express the error structure as:

$$W = \gamma_x X + U, \tag{1.11}$$

where U has mean zero, variance σ_u^2 and it is independent of X. If (X, ϵ, U) are jointly normally distributed, then the regression of Y on W is linear with intercept

$$b_{0*} = b_0 + b_x \mu_x - b_{x*}(\gamma_x \mu_x)$$

and slope

$$b_{x*} = \frac{b_x \gamma_x \sigma_x^2 + \rho_{\epsilon_u} \sigma_\epsilon \sigma_u}{\gamma_x^2 \sigma_x^2 + \sigma_u^2}, \tag{1.12}$$

where $\sigma_{xw} = \gamma_x \sigma_x^2$, $\sigma_w = \gamma_x^2 \sigma_x^2 + \sigma_u^2$ and $\rho_{\epsilon u}$ is the correlation between ϵ and U. Unlike what we said as comment to (1.10), from (1.12) we see that there are two different situations where it is possible that $|b_{x*}| > |b_x|$: if W is biased ($\gamma_x \neq 1$) and if $\rho_{\epsilon u} \neq 0$. In both cases, the measurement error could lead to an opposite effect to the attenuation, so the naive model slope could be less or greater than b_x, depending on the amount of bias and on the correlations between errors in linear regression models (ϵ) and measurement errors (U). To correct the bias we will need most additional information or data. When (1.11) is verified, the residual variance of the linear regression of Y on X is

$$\text{var}(Y|W) = \sigma_\epsilon^2 + \frac{b_x^2 \sigma_x^2 \sigma_u^2 - \rho_{\epsilon u}^2 \sigma_\epsilon^2 \sigma_u^2 - 2 b_x \gamma_x \sigma_x^2 \rho_{\epsilon u} \sigma_\epsilon \sigma_u}{\gamma_1^2 \sigma_x^2 + \sigma_u^2} \ .$$

In Table 1.1, we summarize the primary effects of measurement error in a simple regression model where (Y, X, W) are multivariate normal and the underlying regression model is $Y = b_0 + b_x X + \epsilon$, where X and ϵ are independent and ϵ has mean zero and variance σ^2_ϵ. We present, for each of the four different error models (no error ($W = X$), an unbiased classical-error measurement ($W = X + U$), an unbiased Berkson measurement ($X = W + U$) and a general differential measurement error), the error model squared correlations, intercepts, slopes and residual variances of the linear model relating Y to W.

1.3.4 The Multiple LRM with Measurement Error

We can extend our reasoning to the case with more than one predict variable, among which we will distinguish between explanatory variables measured without error, $\mathbf{Z} = (Z_1, \ldots, Z_m)$, and those that cannot be measure exactly, $\mathbf{X} = (X_1, \ldots, X_p)$.

The linear model is

$$\begin{aligned} Y &= b_0 + \mathbf{b}'_x \mathbf{X} + \mathbf{b}'_z \mathbf{Z} + \epsilon, \\ \mathbf{W} &= \mathbf{X} + \mathbf{U}, \end{aligned}$$

where $\mathbf{X} = (X_1, \ldots, X_p)$ is a matrix $(n \times p)$, with n the number of the observation and p the number of the error-prone covariates, $\mathbf{Z}$ is a matrix containing the m covariates measured without error.

We have to observe immediately that the naive ordinary linear regression does not consistently estimate $(\mathbf{b}'_x, \mathbf{b}'_z)$, but

$$\begin{aligned} \begin{bmatrix} \mathbf{b}_{x*} \\ \mathbf{b}_{z*} \end{bmatrix} &= \begin{bmatrix} \Sigma_{xx} + \Sigma_{uu} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}^{-1} \left\{ \begin{bmatrix} \Sigma_{xy} \\ \Sigma_{zy} \end{bmatrix} + \begin{bmatrix} \Sigma_{u\epsilon} \\ 0 \end{bmatrix} \right\} \\ &= \begin{bmatrix} \Sigma_{xx} + \Sigma_{uu} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}^{-1} \left\{ \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix} \begin{bmatrix} \mathbf{b}_x \\ \mathbf{b}_z \end{bmatrix} + \begin{bmatrix} \Sigma_{u\epsilon} \\ 0 \end{bmatrix} \right\}. \end{aligned} \tag{1.13}$$

It is important to remark that when the model includes both error-prone and error-free covariates, the attenuation due to the effect of a classical

Error-Model	ρ^2_{xw}	Intercept	Slope	Residual Variance
No error	1	b_0	b_x	σ^2_ϵ
Classical	$\frac{\sigma^2_x}{\sigma^2_x+\sigma^2_u}$	$b_0 + b_x\mu_x(1-\rho^2_{xw})$	$b_x\frac{\sigma^2_x}{\sigma^2_x+\sigma^2_u}$	$\sigma^2_\epsilon + b^2_x\sigma^2_x(1-\rho^2_{xw})$
Berkson	$\frac{\sigma^2_x-\sigma^2_{u_b}}{\sigma^2_x}$	b_0	b_x	$\sigma^2_\epsilon + b^2_x\sigma^2_x(1-\rho^2_{xw})$
Differential	ρ^2_{xw}	$b_0 + b_x\mu_x - \frac{b_x\sigma_{xw}+\sigma_{\epsilon w}}{\sigma^2_w}\mu_w$	$\frac{b_x\sigma_{xw}+\sigma_{\epsilon w}}{\sigma^2_w}$	$\sigma^2_\epsilon + b^2_x\sigma^2_x - \frac{(b_x\sigma_{xw}+\sigma_{\epsilon w})^2}{\sigma^2_w}$

Table 1.1. Primary effects of measurement error in simple regression for various types of error models. (Carroll et al., 2006)

measurement error may affect not only the naive estimator of $\mathbf{b}_x$ but also the naive estimator of $\mathbf{b}_z$.

To correct the bias, when $\mathbf{W}$ is unbiased for $\mathbf{X}$ and Σ_{uu} e $\Sigma_{u\epsilon}$ are known or can be estimated, we can use (1.13) to construct a simple method-of-moments estimator:

$$\begin{bmatrix} S_{ww} - \Sigma_{uu} & S_{wz} \\ S_{zw} & S_{zz} \end{bmatrix}^{-1} \begin{bmatrix} S_{wy} - \Sigma_{u\epsilon} \\ S_{zy} \end{bmatrix}$$

where S_{ab} indicates the sample covariance between $\mathbf{A}$ and $\mathbf{b}$. If $\mathbf{W}$ is biased for $\mathbf{X}$ (so $\mathbf{W} = \gamma_0 + \gamma_x \mathbf{X} + \mathbf{U}$), the method-of moments estimator is still useful but we have to employ the error-calibrated variable $\mathbf{W}_* = \hat{\gamma}_x^{-1}(\mathbf{W} - \hat{\gamma}_0)$.

The Nonlinear Regression

With nonlinear measurement error models, we can find qualitative effects similar to those described in the previous paragraph for linear models, obviously more pronounced is the nature of nonlinearity of the model, less relevant are the previous results. An exception is the unbiased Berkson error that, in this context, produces biases, even if, for a comparable ρ_{xw}, they are less severe than that due to classical measurement error. Starting from this consideration, it will give us, in the following Section, the cue to talk about the regression Calibration technique, that estimates an unbiased Berkson predictor by a preliminary calibration analysis to conduct an usual naive analysis with $\widehat{\mathrm{E}[X|W]}$ replacing X.

1.4 The Response Variable Error Model

It is possible that the response variable is affected by measurement error:

$$S = Y + V,$$

where Y is the true response, V is the measurement error and S is the observed response. Some authors (for example Abreyava and Hausman, 2004) think that classical response measurement error can be ignored in regression analysis because it is included into residual error. Thanks to

some simulated results, some authors support that "in linear or nonlinear regressions that have homoscedastic errors about the true line, the only effects of adding unbiased, homoscedastic response measurement error is to increase the variability of the fitted lines and surfaces, and to decrease power for detecting effects. All tests, confidence intervals, etc, are perfectly valid: they are simply less powerful" (Carroll et al., 2006) . We can suppose that Y, without response error, has mean $m_y(\mathbf{Z}, \mathbf{b}_z)$, and variance σ_y^2, while the observed response S has mean $m_y(\mathbf{Z}, \mathbf{b}_z)$ and variance $\sigma_s^2 = \sigma_y^2 + \sigma_v^2$; thus the observed data have the same mean of Y, but a greater variance than the true one.

When we have a nonlinear models (like in this case), we must remember that the inference is often based on linear approximation of the model by Taylor expansion of the parameter $\mathbf{b}_z$ about its true value $\mathbf{b}_{z_0}$, for example

$$y_i = m_y(\mathbf{z}_i, \mathbf{b}_z) + \epsilon_i \quad \approx \quad m_y(\mathbf{z}_i, \mathbf{b}_{z_0}) + f'(\mathbf{z}_i, \mathbf{b}_{z_0})(b_z - \mathbf{b}_{z_0}) + \epsilon_i.$$

The error in the Taylor approximation decreases to zero as $\mathbf{b}_z$ approaches to $\mathbf{b}_{z_0}$. We can observe that when the response variance increase then the approximation becomes less accurate, also the increase of σ_u determines an increasing skewness of $\hat{b}_z$.

1.4.1 Biased Responses

It is possible that the observed response S is biased for the true response Y, in this case we obtain biased estimates for the regression parameters. For example, we consider a model where Y given $\mathbf{Z}$ follows a normal linear model with mean $b_0 + \mathbf{b}_z'\mathbf{Z}$ and variance σ_ϵ^2, if the distribution of S given $(Y, \mathbf{Z})$ follows a normal linear model with mean $\gamma_0 + \gamma_y Y$ and variance σ_v^2, so S is biased and the observed data follow a normal linear model with mean $\gamma_0 + b_0\gamma_y + \gamma_y\mathbf{b}_z'\mathbf{Z}$ and variance $\sigma_v^2 + \gamma_y^2\sigma_\epsilon^2$. If we don't consider the response measurement error, the naive regression estimator does not estimate $\mathbf{b}_z$, but instead $\gamma_y\mathbf{b}_z$. An obvious solution would be to use $\frac{S-\gamma_0}{\gamma_y}$, but often we don't know γ_0 and γ_y, so we have to adopt different strategy to solve the problem (for example see Buonaccorsi, 1996). When it is possible, an useful method

consists in the use of validation data.

Validation Data

If we have some validation data on a simple random subsample of the primary data, we can follow the following procedure at 5 steps to obtain information about (γ_0, γ_y).

1. We use the validation subsample to calculate $\hat{\mathbf{b}}_{z_1}$, that is an estimation of $\mathbf{b}_z$, and (γ_0, γ_y).

2. In the second step we want to obtain an estimated unbiased response $(S - \hat{\gamma}_0)/\hat{\gamma}_y$ that we use in another iteration of the analysis to get a second estimate $\hat{\mathbf{b}}_{z_2}$.

3. Now we can use bootstrap techniques to estimate the joint covariance matrix of these estimates (Σ).

4. In this step we calculate the best weighted combination of $\hat{\mathbf{b}}_{z_1}$ and $\hat{\mathbf{b}}_{z_2}$

$$\hat{\mathbf{b}}_z = (\mathbf{J}'\Sigma^{-1}\mathbf{J})^{-1}\mathbf{J}'\Sigma^{-1}(\hat{\mathbf{b}'}_{z_1}, \hat{\mathbf{b}'}_{z_2}),$$

   where $\mathbf{J} = (\mathbf{I}, \mathbf{J})$, $\mathbf{J}$ is the identity matrix $(r \times r)$ and r is the number of elements in $\mathbf{b}_z$.

5. At last, we can use $(\mathbf{J}'\hat{\Sigma}^{-1}\mathbf{J})^{-1}$ as estimated covariance matrix for the estimates $\hat{\mathbf{b}}_z$.

Often it is quite difficult and expensive to construct a validation dataset, so other methods have been implemented; for example it is possible to obtain information by replicated measurement of S or using external data in which Y is observed.

1.4.2 The Logistic Regression Model for Discrete Data

As we have mentioned at the beginning, with discrete data the response error corresponds to misclassification: there is an error when $Y = a$ and $S = b$, with a and b different modes of the categorical response variable.

We consider a logistic regression model with response probability

$$Pr(Y = 1|\mathbf{Z}) = H(b_0 + \mathbf{b}_z'\mathbf{Z}), \tag{1.14}$$

where $H(\cdot)$ is the logistic distribution function, and we suppose that the misclassification error is independent of $\mathbf{Z}$ and the probabilities of correct classification are:

$$\begin{aligned} Pr(S = 1|Y = 1, \mathbf{Z}) &= \pi_1 \\ Pr(S = 0|Y = 0, \mathbf{Z}) &= \pi_0 \ . \end{aligned}$$

Because of misclassification errors, the observed data don't follow the logistic model (1.14), but the more complex form

$$Pr(S = 1|\mathbf{Z}) = (1 - \pi_0) + (\pi_1 + \pi_0 - 1)H(b_0 + \mathbf{b}_z'\mathbf{Z}) \ , \tag{1.15}$$

so the response misclassification is considerably biased for the true probability response. This problem has been studied from many authors: Paulino et al. (2003) studied the binomial regression with misclassification, Prescott and Garthwaite (2002) focused on validation substudies, Neuhaus (2002) developed the theme of analysis of clustered and longitudinal binary data subject, and Ramalho (2002) has addressed the problem of choice-based samples.

In practical analysis, the difficulty of estimation of the underlying risk function significantly affecting the ability of data of the data to identify the classification probabilities. Copas (1988) states that "accurate estimation of the misclassification parameters is very difficult if not impossible unless n is extremely large", it is possible, at least, to conduct a sensitivity analysis for plausible values of the misclassification probabilities.

To estimate $(\pi_0, \pi_1, b_0, \mathbf{b}_z')$, if we assume that the misclassification probabilities are independent of $\mathbf{Z}$, we can using maximum likelihood or Bayesian methods. We indicate the probability model (1.15) as $\Psi(S, \mathbf{Z}, \pi_0, \pi_1, b_0, \mathbf{b}_z')$ and we maximize, for example with EM-algorithm, the

loglikelihood function

$$\sum_{i=1}^{n}\{s_i log[\Psi(S, \mathbf{Z}, \pi_0, \pi_1, b_0, \mathbf{b}_z')] + (1 - s_i) log[1 - \Psi(S, \mathbf{Z}, \pi_0, \pi_1, b_0, \mathbf{b}_z')]\}. \tag{1.16}$$

Even if it is not a easy estimation, sometimes we have the possibility to know the value of true Y for a subsample of the data, so we can directly estimate the classification probabilities maximizing the loglikelihood function above in $(b_0, \mathbf{b}_z')$.

When we can have validation data, we can apply *pseudolikelihood* methods that consist in estimating π_1 (π_0) as the fraction of the subject, in the validation study, who are correctly classified among those whose true value is $Y = 1$ ($Y = 0$), assuming that these are known, and then maximize the likelihood (1.16). In order to apply these estimation methods it is important that the selection into the validation study does not depend on $\mathbf{Z}$ or on the observed value S, otherwise these procedures can cause biased estimations.

1.5 The Regression Calibration Method

Regression calibration is well known approximate model to correct biases in measurement error regression models and it is widely used in epidemiological field; we assume the existence of a calibration substudy, where accurate and crude measurement methods are related by a second regression analysis. As we have seen in the previous paragraphs, if we have a classical error model, where only one variable is measured with error, the regression coefficient estimate b_x using traditional regression analysis is biased toward the null value; where there are two or more error-in variables, biases may point in either direction, and there are important cases in which bias points away from the null. Moreover, estimates for variables that are not measured with error will also be biased (Rosner et al., 1990). Regression calibration for linear models will correct such biases if the assumptions are met. The key of this method is the substitution of $\mathbf{X}$ with the expected value of the true variable

conditional on observed data measured (**Z,W**). Regression calibration is quite simple and, if the approximation is sufficiently accurate, it is applicable to any regression model. The algorithm consists of three phases:

1. First we have to estimate the regression of **X** on (**Z,W**), $m_{\mathbf{X}}(\mathbf{Z}, \mathbf{W}, \gamma)$ that depend on parameters γ, estimated by $\hat{\gamma}$;

2. then we can run standard analysis to obtain parameter estimates, substituting the unobserved **X** by its estimate;

3. finally we must adjust the resulting standard errors to account for the estimation of γ.

To better understand this procedure, we imagine a model where

$$E[\mathbf{Y}|\mathbf{Z,W}] = m_{\mathbf{Y}}[\mathbf{Z}, \mathbf{X}, B],$$

for some unknown parameter B. If we substitute, in this expected value, **X** with its estimated value, we have a modified model for observed data:

$$E[\mathbf{Y}|\mathbf{Z,W}] \approx m_{\mathbf{Y}}[\mathbf{Z}, m_{\mathbf{X}}(\mathbf{Z}, \mathbf{W}, \gamma), B].$$

The replacement of **X** with its regression on (**Z,W**) is not always an easy passage of the algorithm, except in same simple cases, for example if we have validation data, we can regress **X** on the other covariates (**Z,W**) in the validation subsample. In other cases we can have an unbiased instrument **T** for a subset of the original sample, so the regression of **T** on (**Z,W**) is the same as the regression of **X** on (**Z,W**).

In practical applications, in addition to the difficulty of estimating the regression of **X** on (**Z,W**), this method has two drawbacks: it doesn't exist a simple approximation that can always be accurate and it shows some difficulties with high non linear models.

Here we briefly present some considerations of this method following the scheme proposed by Fraser and Stram (2001). With the regression calibration

we often have a loss of statistical power as compared with the case where we know X. One reason for this phenomenon is due to the reduced variability of the predictor variable in the new regression model, and usually this leads to a reduced power to detect a nonzero slope in the regression analysis: $\mathrm{E}[\mathbf{X}|\mathbf{W},\mathbf{Z}] = \mathbf{X}_{RC}$, that is the predictor that we use instead of the unknown $\mathbf{X}$, is less variable than $\mathbf{X}$, as we can see from

$$\mathrm{Var}[\mathbf{X}] = \mathrm{Var}\{\mathrm{E}[\mathbf{X}|\mathbf{W},\mathbf{Z}]\} + \mathrm{E}\{\mathrm{Var}[\mathbf{X}|\mathbf{W},\mathbf{Z}]\}.$$

We can identify another cause of the reduction of statistical power in the greater variability of $\mathbf{Y}$ given $\mathbf{W}$ than in $\mathbf{Y}$ given $\mathbf{X}$. However, as we will see in the following Sections, for binary $\mathbf{Y}$, this second issue can be neglected, since the variance of $\mathbf{Y}$ is uniquely determined by its mean value.

When we have an univariate linear calibration equation, we can verify that

$$\mathrm{Var}\{\mathrm{E}[\mathbf{X}|\mathbf{W}]\} = R^2\mathrm{Var}[\mathbf{X}], \tag{1.17}$$

where R is the correlation coefficient between $\mathbf{X}$ and $\mathbf{W}$. From 1.17, we can deduce that, when we use regression calibration method, we need $\frac{1}{R^2}$ more units to detect a nonzero regression between $\mathbf{Y}$ and $\mathbf{W}$ compared to the case where $\mathbf{X}$ is available. In summary, when we apply a regression calibration model, the statistical power may be adversely affected by different factors:

- small sample size;
- multicollinearity between the covariates;
- poor validity, as measured by the correlation between $\mathbf{X}$ and $\mathbf{W}$ (this depends in part on the residual error in the calibration equations);
- imprecision in the estimate of $E[\mathbf{X}|\mathbf{W},\mathbf{Z}]$ due to imprecision of the calibration equation coefficients.

For the asymptotic formulae of the standard errors, we refer to studies of Carroll and Stefanski (1990).

1.6 The Simulation Extrapolation Method

The simulation and extrapolation method (SIMEX), initially proposed by Cook and Stefanski (1994) and subsequently developed by several authors, has become a useful tool for correcting estimates in the presence of additive measurement error, and more generally to all cases in which the measurement error generating process can be simulated through Monte Carlo techniques. In 2006, Küchenhoff et al. (2006) applied the same basic idea of simulation and extrapolation to the case of misclassification and he proposed the model MCSIMEX. It is a functional method since there are no assumptions made about the distribution of the unobserved true covariate. Intuitively, the idea of this method is to generate new observations with a greater measurement error as the original sample, to determine a trend of the bias due to measurement error versus the variance of the added measurement error, and to extrapolate this trend back to the case of no measurement error. We will use the following function

$$G(\sigma_u^2) := \sigma_u^2 \to b^*(\sigma_u^2),$$

where b^* is the limit to which the naive estimator converges as the sample size $n \to \infty$. In fact, it is quite simple to see that $G(0) = b$ is the true parameter, and $G(\sigma_u^2) = \hat{b}_{x*}$ is the ordinary least square estimator (1.6), also called naive estimator. We recall that $\hat{b}_{x*}$ consistently estimates $b_x\lambda$ and it is biased for b_x if $\sigma_u^2 > 0$.

1.6.1 SIMEX in Simple Linear Regression

To describe the basic idea of SIMEX for a simple linear regression model, we start from

$$Y = b_0 + b_x X + \epsilon,$$

with additive measurement error $W = X + U$; $U \sim D(0, \sigma_u^2)$, where D is any unknown distribution, is independent of (Y, X).

We want to show that the effect of measurement error on the estimation

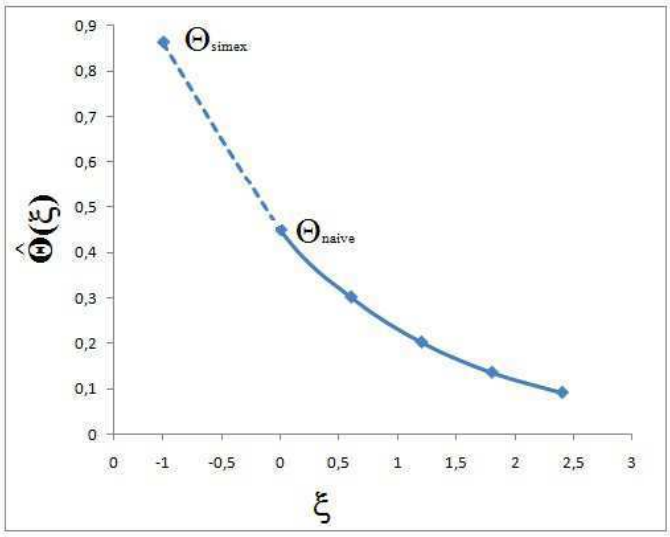


Figure 1.1. A generic SIMEX plot showing the effect on a statistic of adding measurement error with variance $\xi\sigma_u^2$ to the data when estimating a generic parameter ξ (Carroll et al., 2006)

of a generic parameter, Θ, can be defined experimentally via simulation. We suppose to have, in addition to the original data used to calculate $\hat{b}_{x*}$, $M-1$ other datasets created with increasing measurement error variance, so the m-th dataset has $(1+\xi_m)\sigma_u^2$, where $0=\xi_1<\xi_2<\cdots<\xi_M$ are known.

In Figure 1.1, we can see a graphical representation of the Simulation Extrapolation procedure. In the abscissa we have ξ, in the ordinate we find the estimated parameter $\hat{\Theta}$. When $\xi=0$ we have the naive estimate, while the SIMEX estimate is an extrapolation of the estimates $\hat{\Theta}$ to the case where $\xi=-1$. Obviously we can not really add measurement errors with negative variance to data, but the purpose of this method is just adding pseudo error with positive variance that we allow the estimation of the form of the bias as a function of ξ. In this way, we can extrapolate the case, purely hypothetical, of adding pseudo-errors with negative variance.

We indicate with $\hat{b}_{x,m}$ the least squares slope estimate, obtained by the m-th dataset; it consistently estimates

$$b_x\frac{\sigma_x^2}{\sigma_x^2+(1+\xi_m)\sigma_u^2}.$$

To solve this system, where the dependent variable is $\hat{b}_{x,m}$ and the independent variable is ξ_m, we set the analysis as a nonlinear regression model with data $\{(\xi_m, \hat{b}_{x,m}), m = 1, \ldots, M\}$. Asymptotically the mean function of this regression is

$$E(\hat{b}_{x,m}|\xi) = G(\xi) = b_x \frac{\sigma_u^2}{\sigma_x^2 + (1+\xi)\sigma_u^2}, \qquad \xi \geq 0.$$

Now we can obtain the parameter of interest from $G(\xi)$, because if we extrapolate back this function to the case of no measurement error, that corresponds to $(1+\xi_m)\sigma_u^2 = 0$, $\xi_m = -1$, we obtain $G(-1) = b_x$.

We can summarize the SIMEX algorithm in five steps (Hardin et al., 2003):

- The *Preliminary fitting step*, where we fit the model to estimate $\hat{b}_{x,m}$ and the measurement error variance $\hat{\sigma}_u$.
- In the *Simulation step*, we add to the original data, $M-1$ datasets generated with successively larger measurement error variances $\xi_m\sigma_u$; for the m^{th} dataset, the total measurement error variance is $\sigma_u + \xi_m\sigma_u = (1+\xi_m)\sigma_u$.
- During the *Estimation step*, we obtain estimates from each of the m^{th} generated contaminated data sets, using an algorithm that would have been used if there were no measurement error datasets.
- We repeat simulation and estimation step D times (D should be very large) and we calculate the average estimated parameter for each of the M level of error disturbance used. Then we plot these mean values against the ξ_m values and we use a regression method to fit an extrapolant function to the averaged, error-contaminated estimates.
- In the *Extrapolation step*, finally we extrapolate to the ideal case of no measurement error ($\xi = -1$).

In literature have been proposed many extrapolant functions, we remember three of the most used:

- *Linear extrapolant*

$$x^* = b_0 + b_1\xi,$$

it is the simplest one and it is useful when the error is small and the extrapolant function is close to linear. x^* is the mean of simulations runs.

- *Rational linear extrapolant*

$$x^* = b_0 + \frac{b_1}{b_2 + \xi},$$

it reproduces the usual method-of-moments estimators for the case of multiple linear regression with non-IID errors. It can produce instable estimation, especially if the measurement error effects on a parameter are negligible ad a nearly constant extrapolant is required (for example, when we estimate the coefficient of error-free covariate uncorrelated with W).

- *Quadratic extrapolant*

$$x^* = b_0 + b_1\xi + b_2\xi^2,$$

it is a conservative corrections for attenuations and it is usually numerically stable.

It is quite simple to generalize all that we have said above to the multivariate case (1.13), where $\mathbf{W} = \mathbf{X} + \mathbf{U}$ with $\mathbf{U} \sim \mathrm{N}(0, \Sigma_u)$.

Chapter 2

Latent Models for Social Measurement

2.1 Introduction

Almost no statistical problem is characterized by one variable, the phenomena of interest are often the result of multiple concurrent elements that are difficult to control conjointly. When we want to analyze simultaneously these multiple characters, we refer to multivariate statistical methods, that are not only easy generalization of univariate statistical procedures because of the complexity increases exponentially with increasing the size of the problem.

In social studies, we often have to analyze multidimensional aspects, such as the job satisfaction, that are not directly observable or measurable through traditional survey instruments, and we want to define a scientific measurement for them, taking into account all the existing links between the different aspects involved (Allen, 1979; Bartholomew, 1996; Skrondal and Rabe-Hesketh, 2004). These concepts are defined as constructs or latent factor: i.e latent variables that are not directly observable but that can be inferred (through a mathematical model) from other variables that are observed (directly measured).

In this chapter, we will present two different approaches to the

categorical data analysis (Section 2.2): the Item Response Theory and the Underlying Variable Approach. In Section 2.4, we will deepen the Item Response Models, presenting the basic Rasch Model for dichotomous data and some polytomous item response models (the Graded Response Model, the Partial Credit Model and the Rating Scale Model). Finally, we will do some reflection on the reliability of measures obtained with these models (Section 2.4.4). The Underlying Variable Approach will be treated in Section 2.3, where we will introduce the Structural Equation Model. After having briefly presented some essential features of them, we will focus on two particular models, the Muthén SEM (Section 2.3.3) and the Reticular Action Model (Section 2.3.5).

2.2 Categorical Data Analysis and Measurement

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. Objectivity requires that the measure is assigned to construct independently from the observer; for the reproducibility, instead, the resulting comparison between objects should be invariant, other conditions being equal.

Over the years many authors have studied this issue thoroughly proposing very different solutions, especially related to their specific scopes. To summarize, we can identify two main approaches to analyzing multivariate latent aspects taking into account the categorical nature of the observed variables (Cagnone et al., 2010):

- *Underlying Variable Approach* (UVA);
- *Item Response Theory* (IRT).

2.2.1 Underlying Variable Approach

The Underlying Variable Approach, or UVA, (Jöreskog, 1973 and 1979; Muthén, 1984) assumes that the observed categorical outcomes are incomplete observations of unobserved continuous variables: underlying each of the categorically observed variables Y_j's (that has c_j categories) there is a continuous variable Y_j^* which is actually measuring the underlying latent factors θ, not directly observable. We assume the traditional linear factor analysis model for the "partially observed" variables Y_j^*:

$$Y_j^* = \Lambda\theta + \epsilon.$$

For categorical observed variables we assume

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

with the threshold values $-\infty < \tau_{j,1} < \tau_{j,2} < \ldots < \tau_{j,c_j-1} < \infty$.

In literature several methods for fitting this model have been proposed, in Section 2.3.3 we will present the Muthén model (1984), implemented in the M*plus* software. It is a three step approach model, in the first step we estimate thresholds based on univariate distributions, then we estimate polychoric correlations, and in the last step we use them as input into a usual factor analysis routine.

2.2.2 Item Response Theory Approach

With the Item Response Theory, or IRT, approach, the observed variables are treated as they are (Rasch, 1960 and 1961; Andrich, 1978; Baker, 1985; Baker and Kim, 2004). The unit of analysis is the entire response pattern of a subject, so we have no loss of information. For a given response vector $\mathbf{y}$,

we can write its distribution as

$$p(\mathbf{y}) = \int p(\mathbf{y}, \theta) df = \int p(\mathbf{y}|\theta) p(\theta) df.$$

We assume local conditional independence for the elements y_j in $\mathbf{y}$ given θ, so that

$$p(\mathbf{y}|\theta) = \prod_{j=1}^{p} p(y_j|\theta).$$

Often, we assume that the underlying factors θ are distributed as $\mathrm{N}(0, I)$.

In this context, we consider the binary case where each y_j is equal to 0 or 1; in the following paragraphs we will extend it to the ordinal case. A natural distribution to choose for $p(y_j|\eta)$ is the logistic function,

$$p(y_j|\theta) = \pi_j(\theta) = \frac{exp[\alpha_j(\theta_i - \delta_j)]}{1 + exp[\alpha_j(\theta_i - \delta_j)]},$$

where δ_j is the difficulty of item j, and the vector of discrimination parameters, α_j, represent the slopes relating each of the m factors to the j-th item. We could use other link functions besides the logistic, for example the normal ogive.

In literature, is even open the debate on which of the two approaches, IRT or UVA, is better. There are some studies (for example Jöreskog and Moustaki, 2001 and Huber et al., 2004) showing that IRT approach has better properties (for example in term of estimates accuracy and model fit), probably because it is a full information approach, while UVA is based on limited information estimation methods. However, the IRT approach is computationally heavy, the likelihood is not available in closed form and it is approximated by numerical integration requiring several quadrature points. With UVA the computer time does not increase with sample size and it is feasible for many variables and many factors, using bivariate moments. A rule of thumb could be to choose UVA when we want to identify clusters of variable measuring the same factors; at the contrary if we have a set of items measuring only one factor and we use them to

specifically score and rank individuals, the IRT approach is more appropriate.

Our work arises from these various considerations on the two previously proposed approaches. As mentioned earlier, in the real applications analyst are often forced to choose which of them to use: we want to compare, on an equal databases, the results obtained by applying both UVA and IRT, their fitting and other proprieties of the obtained estimates. The idea of how to make this comparison, it was suggested by a work of Gibbons et al. (2007) where they indirectly showed, proposing their bifactorial model, how to implement an IRT model with a Structural Equation Model (that is an Underlying Variable Model). The details of this implementation are showed in their article and some of them will be included in the third chapter.

2.3 Structural Equation Models (SEMs)

In Section 2.2, have seen that exists a different approach for a categorical data analysis, different from IRT: the Underlying Variable Approach (UVA). One of the most used family of models that belongs to this framework is the Structural Equation Model (SEM). In literature, there are other similar methods to those described in this chapter: Covariance Structural Analysis, Covariance Structural Model or Causal Modeling, that differ from SEM for the analytical approach and/or for the correlation structure used in the estimation process.

The origins of these methods can be found in the early years of the last century with the development of exploratory factor analysis, usually credited to Spearman (1904), and the basics of path analysis developed by Wright (1934). These measurement (factor analysis) and structural (path analysis) approaches were integrated in the early 1970s by Jöreskog (1973), Keesling (1972), and Wiley (1973), into a framework that Bentler and Yuan (1999) called the JKW model. In the 90's, SEM methods expanded in many area, such as genetics, sports medicine, development psychology, public health and education.

The widespread use of these models made difficult to identify a single

pattern common to all of them, it is possible, however, to define some common features:

1. SEM requires that the researcher specify a *priori* model, not only for confirmatory analysis, but also for exploratory ones. He has to decide which variables are assumed to affect the others and the directionality of these effects; these hypotheses will be evaluated in the analysis.

2. With SEM, we make an explicit distinction between observed and latent variables.

3. Historically, the basic statistic is the covariance, that represents the strength of the association between two variables; actually, SEM is a very flexible analytical approach that can incorporate also between-group and within-group mean comparisons, as in a standard ANOVA.

4. We can apply SEM methods to experimental or nonexperimental data, or to a mix of them.

5. The SEM family includes many standard procedures. For example, ANOVA is a special case of multiple regression, both these procedures are members, with exploratory factor analysis, of general linear model (GLM), and GLM are special instances of SEM.

6. SEM is a large-sample method. The sample size is affected by different factors, for example the model complexity, the required result stability, the type of estimation algorithm. Referring to the literature (MacCallum, 2000), we can say that for descriptives purposes, sample size less than 100 will be considered small, between 100 and 200 medium and more than 200 large.

7. Usually, in the SEM literature we find unstandardized estimation, not only because the most widely estimation methods used assume the analysis of unstandardized variables, but also because there are situations where important information will be lost when variable are standardized: for example the analysis of SEM across multiple

samples that differ in their variabilities or longitudinal measurement of variables that show increasing (or decreasing) variability over time.

2.3.1 The SEM Approach

Even though they are sometimes very different, for the SEM approach we can identify the following six basic steps:

1. *Specification of the model.* The researcher must specify his hypotheses (about the relations among the latent and the observed variables) in the form of a structural equation model. It is possible to describe the model as a series of equation or with a drawn diagram (we will present the notation and the graphic representations in the following paragraphs).

2. *Model identification conditions*, that is if it is theoretically possible to derive a unique estimate of every model parameter. In this step we want to avoid multicollinearity and singular matrices.

3. *Collection, preparation and screening* of the data.

4. *Model estimation.* In this step we evaluate the model fit to determine how well the model explains the data, we interpret the parameter estimates for specific effects, and we consider some equivalents models that explain the data just as well as our preferred model but with a different configuration of hypothesized relations.

5. If necessary, we *respecify the model* and we execute again the step 2, 3 and 4 till we obtain satisfactory results.

6. Finally, we *accurately and completely describe the analysis* in written reports.

Although many authors cover their analysis with these 6 steps, in reality there are still two very important steps:

7. *Replication of the results*, estimating the models across independent samples.

8. *Apply the results.*

As it happens often in recent literature, there is no uniformity of notation with respect to the SEM; the most widely used symbology is probably the one used in the LISREL context, with Greek letters and matrix algebra.

2.3.2 Equations Notation and Graphic Representations

In this Section, we want to briefly review some specification issues to establish a common terminology for the discussion that follows. We refer to the classic LISREL notation (Bollen, 1989), even if it could be some little modification to uniform the notation with IRMs and other models proposed in this work.

A full structural equation model can be split in two submodels:

- the Structural Model is the set of latent variables in the model, together with the direct effects connecting them;
- the Measurement Model is that part (possibly all) of a SEM model which deals with the latent variables and their indicators.

Formally we have:

The **Structural Model**

$$\theta = \mathbf{B}\theta + \zeta \tag{2.1}$$

The **Measurement Model**

$$\mathbf{y} = \Lambda^y\theta + \epsilon \tag{2.2}$$

$$\mathbf{x} = \Lambda^x\theta + \delta \tag{2.3}$$

Equation (2.1) is called the latent variable (or structural) model and expresses the hypothesized relationships among the constructs. θ is a vector $(m \times 1)$ that contains the latent constructs, endogenous or exogenous (in

standard LISREL notation endogenous and exogenous have different symbols, but for our analysis it has no relevance to separate them). **B** is the coefficient matrix for the effects of constructs on each other. A structural coefficient is the measure of the amount of change in the effect variable expected given a one unit change in the causal variable and no change in any other variable. ζ is the vector $(m \times 1)$ of disturbances, and it represents errors in structural equations. When we specify that latent variables have simultaneous effects on each other (so that the **B** matrix has nonzero elements both above and below the diagonal) and/or if errors in equations are allowed to be correlated, the SEM is called nonrecursive. We have to pay attention with these models, because there can be some problem of model identification, stability of reciprocal effects, and interpretation of measures of variation accounted for in endogenous constructs (Schaubroeck, 1990; Teel et al., 1986). If **B** is subdiagonal and the ζ are uncorrelated, the model is said to be recursive.

Equations (2.2) and (2.3) represent the measurement model which links the constructs to observable indicators. The vectors $\mathbf{y}$, $(p \times 1)$, and $\mathbf{x}$, $(q \times 1)$, contains the measures of the p endogenous constructs, and the q exogenous indicators, respectively. Indicators are observed variables, sometimes called manifest variables or reference variables, such as items in a survey instrument. Four or more is recommended and three is acceptable and common practice. Λ^y and Λ^x are the coefficient matrices for the relations of $\mathbf{y}$ and $\mathbf{x}$ with θ (they are also called loadings). The vectors ϵ and δ are the disturbances and they represent the errors in variables or measurement errors. Whereas regression models implicitly assume zero measurement error (and so, as we have seen in the first chapter, to the extent such error exists, regression coefficients are attenuated), error terms are explicitly modeled in SEM and as a result SEM estimators are unbiased by error terms, whereas regression coefficients estimators are not. Generally (but not always) the measurement model possesses simple structure such that each observed variable is related to a single latent variable. Models with simple structure and no correlated measurement errors represent unidimensional construct measurement, which is frequently considered to be a highly desirable characteristic of measurement

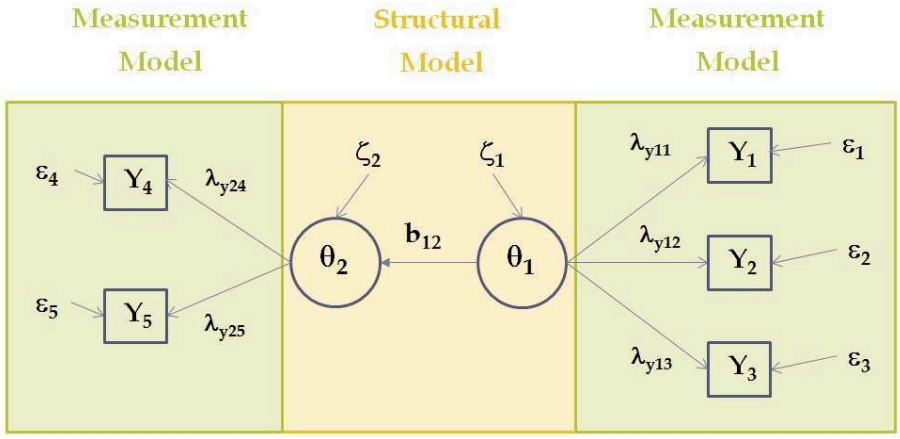


Figure 2.1. Structural Equation Model path diagram

(Hattie, 1985; Anderson and Gerbing, 1988).

As we have said before, a SEM is often represented with a path diagram. The Figure (2.1) shows a structural equation model for two latent factors, θ_1 is measured by three indicators (Y_1, Y_2, Y_3) and θ_2 by two indicators (Y_4 and Y_5); furthermore θ_2 has a dependent relationship with θ_1. A SEM diagram commonly has certain standard elements: latent variables are circles, indicators are rectangles, error and residual terms are not included in a geometrical form, single-headed arrows are causal relations (note causality goes from a latent to its indicators). Coefficient values may be placed on the arrows from latent variables to indicators (the loadings $\lambda_{11}, \ldots, \lambda_{25}$), or from one latent to another (the structural coefficient b_{12}). Each latent variable has an error term (ζ_1, ζ_2), sometimes called a disturbance term or residual error, not to be confused with the indicator error $(\epsilon_1, \epsilon_2, \ldots, \epsilon_5)$ associated with the observed variables.

The extended formulation for the model in Figure (2.1) is:

$$\begin{aligned}
\theta_1 &= \zeta_1 \\
\theta_2 &= \beta_{12}\theta_1 + \zeta_2
\end{aligned}$$

$$\begin{aligned}
Y_1 &= \lambda_{11}\theta_1 + \epsilon_1 \\
Y_2 &= \lambda_{12}\theta_1 + \epsilon_2 \\
Y_3 &= \lambda_{13}\theta_1 + \epsilon_3 \\
Y_4 &= \lambda_{24}\theta_2 + \epsilon_4 \\
Y_5 &= \lambda_{25}\theta_2 + \epsilon_5.
\end{aligned}$$

2.3.3 The Muthén SEM

Bengt Muthén is an important and famous statistician that, since the 70's, explored the research theme of SEM. He implemented a specific software (M*plus*; Muthén and Muthén, 2007) that provides researchers with a flexible tool to analyze, between the different models available, also the SEM. In this paragraph, we consider the specification and estimation of SEMs with latent variables having multiple indicators, not all of which are continuous. We consider a linear structure for continuous latent variables; while, in the measurement part, we could have dichotomous and ordered polytomous observed variables and/or continuous indicators. The model that we present, and that we have used in the simulation study presented in chapter 3, are drawn from several works of Muthén (1979, 1981, 1983, 1984). The author, obviously, refers to the Jöreskog-Sörbom (1973) methodology ("LISREL") for structural equation models to handle properly categorical indicators in addition to continuous ones. He makes a distinction between models with observed independent variables and models without them, but we consider only the second ones because, as we have already said in the previous paragraphs, they are more closely to IRMs that we have implemented in simulation study.

We consider the vector θ, $(m \times 1)$, of continuous latent variable constructs,

and we observe the vector $\mathbf{y}$, $(p \times 1)$, which consist of both dependent and independent variables (in a structural equation modeling sense), that could be ordered categorical or continuous; we indicate with $\mathbf{y}^*$, the $(p \times 1)$ vector of continuous latent response variables associated at the observed variables. For an ordered polytomous y_j with c_j categories ($c_j \geq 2$), we have

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

If y_j is continuous, the latent response variable is directly observed

$$y_j = y_j^* \ .$$

Now we can present the linear measurement structure, that follows the structure of the measurement model presented in (2.2)

$$\mathbf{y}^* = \nu + \Lambda\theta + \epsilon$$

and the linear structural equation system, that follows the structure of the structural model presented in (2.1)

$$\theta = \alpha + \mathbf{B}\theta + \zeta,$$

where $\mathbf{B}$ has zero diagonal elements and $(\mathbf{I} - \mathbf{B})$ is non-singular. Muthén assumes a multivariate normal distribution for Y^*. Due to the normality specification, it suffices to consider first and second-order moments for the latent response variables.

$$\mathrm{E}[Y^*] = \nu + \Lambda(\mathbf{I} - \mathbf{B})^{-1}\alpha + \Lambda(\mathbf{I} - \mathbf{B})^{-1}\Gamma X \tag{2.4}$$

$$\mathrm{Var}[Y^*] = \Lambda(\mathbf{I} - \mathbf{B})^{-1}\Psi(\mathbf{I} - \mathbf{B})'^{-1} + \Xi, \tag{2.5}$$

where Ξ is the covariance matrix of ϵ.

With categorical Y variables, the distribution of the observed variables is deduced by integrating over the corresponding latent response variables Y^* (Muthén, 1979; Muthén and Christoffersson, 1981).

2.3.4 Estimation Procedure for the Muthén SEM

Referring to the model presented in (2.4) and (2.5), we now present the Muthén estimation approach for SEM, which uses weighted least squares with limited, first and second order sample information. This is a three stages estimation approach:

- First, we estimate first order statistics by maximum-likelihood (ML);
- then, we estimate second order statistics by conditional ML for given first stage estimates;
- in the third stage, which is common to all situations, the model parameters will be consistently estimated, using the first and second order statistics generated by the previous stages.

The details of the first two stages will vary depending on the type of indicators involved and they could found in the aforementioned works of Muthén.

Fist of all, we summarize the structure of the general Muthén SEM from (2.4) and (2.5) in two parts, and we consider the vectors σ_1 and σ_3 (σ_2 is used by Muthén to identify the part corresponding the observed independent variables $\mathbf{x}$ in equation (2.3), that we are not considering in this context):

- Part 1: the mean/threshold regression intercept structure

$$\sigma_1 = \Delta^* \mathbf{K}_\tau \tau - \mathbf{K}_\nu [\nu + \Lambda (\mathbf{I} - \mathbf{B})^{-1} \alpha],$$

- Part 3: the covariance/correlation residual correlation structure

$$\sigma_3 = \mathbf{K} \times \mathrm{vec}\{\Delta [\Lambda (\mathbf{I} - \mathbf{B})^{-1} \Psi (\mathbf{I} - \mathbf{B})'^{-1} \Lambda' + \Xi] \Delta\}.$$

Δ is a diagonal matrix of scaling factors useful in multiple-group analysis with categorical variables and Δ^* contains the same element as Δ but diagonal elements are duplicated for categorical variables with more than one threshold (more than two categories). Similarly, $\mathbf{K}_\tau$ and $\mathbf{K}_\nu$ distribute elements from the vectors they pre-multiply, where $\mathbf{K}_\tau$ has a row of zeros for each continuous variable, and $\mathbf{K}$ selects lower-triangular elements from the vector of symmetric matrix elements it pre-multiplies, where diagonal elements are only included if the corresponding observed variable is continuous. The vec operator creates a column vector from a matrix by stacking its the column vectors below one another. Part 1 contains the intercept structure, i.e. the first two terms of (2.4), while Part 3 contains the residual covariance/correlation structure (2.5).

In the first and second stage we consistently estimate by ML the elements of σ_1 and σ_3 as $\mathbf{s}_1$ and $\mathbf{s}_3$; in the third stage, we estimate the model parameters minimizing the weighted least squares fitting function

$$F_3 = (\mathbf{s} - \sigma)' \mathbf{W}^{-1} (\mathbf{s} - \sigma),$$

where $\mathbf{s}' = (\mathbf{s}_1', \mathbf{s}_2')$ and $\mathbf{W}$ is a positive definite weight matrix. All the analytical details of the estimation procedure can be found in the aforementioned work of Muthén, in particular in Mutén (1984).

2.3.5 The Reticular Action Model (RAM)

For the simulation study, in addition to Muthén approach to SEM, we used the Fox (2006) approach, implemented in the *R* package "*sem*". Fox refers to the Reticular Action Model (RAM) formulation, proposed by McArdle (McArdle 1980; McArdle and McDonald, 1984), with the notation presented by McDonald and Hartmann (1992).

The fundamental equation is

$$\mathbf{V} = \mathbf{A}\mathbf{V} + \mathbf{U},$$

where

- **V** contains the indicator variables (Ys), directly observed exogenous variables (**X**), and the latent variables in the model (η);
- **U** contains directly observed exogenous variables, measurement-error variables (ε), and structural disturbances (ζ);
- the matrix **A** includes structural coefficients (β) and factor loadings (λ); typically it is a sparse matrix, containing many 0's .

As we have already explained in paragraph 2.3.3, we will not consider observed exogenous variables. We have m latent factors each with p_h $(h = 1, \ldots, m)$ indicator variables $(p = p_1 + p_2 + \cdots + p_m)$.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_p \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & \lambda_{11} & \cdots & \lambda_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_{p1} & \cdots & \lambda_{pm} \\ 0 & \cdots & 0 & \beta_{11} & \cdots & \beta_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \beta_{m1} & \cdots & \beta_{mm} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_p \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \\ \zeta_1 \\ \vdots \\ \zeta_m \end{bmatrix}$$

The final component of the RAM formulation is **P**, the covariance matrix of **U**. Assuming that all of the error variables have expectations of 0, and that all other variables have been expressed as deviations from their expectations, $\mathbf{P} = \mathrm{E}(\mathbf{U}\mathbf{U}')$:

$$\mathbf{P} = \mathrm{Var} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \\ \zeta_1 \\ \vdots \\ \zeta_m \end{bmatrix} = \begin{bmatrix} \xi^{\epsilon}_{11} & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \xi^{\epsilon}_{22} & 0 & \cdots & 1 \\ 0 & \cdots & 0 & \psi_{11} & \cdots & \psi_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \psi_{12} & \cdots & \psi_{mm} \end{bmatrix},$$

where ξ^{ϵ}_{11} is the variance of ε_1, ψ_{11} is the variance of ζ_1 and ψ_{1m} is the covariance between ζ_1 and ζ_m. **P** is a $(l \times l)$ matrix, where $l = m + p$ is the

number off variables in $\mathbf{U}$.

The key to estimating the model is the connection between the covariances of the observed variables, which may be estimated directly from sample data, and the parameters in $\mathbf{A}$ and $\mathbf{P}$. To pick out the observed variables, we define the $(p \times l)$ selection matrix $\mathbf{J}$

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \end{bmatrix},$$

where $\mathbf{I}_p$ is the $(p \times p)$ identity matrix and $\mathbf{0}$ is a $(p \times m)$ matrix of 0.

The model implies the following covariances among the observed variables:

$$\mathbf{C} = \mathrm{E}(\mathbf{JVV'J'}) = \mathbf{J}(\mathbf{I}_l - \mathbf{A})^{-1}\mathbf{P}[(\mathbf{I}_l - \mathbf{A})^{-1}]'\mathbf{J'}.$$

Let $\mathbf{S}$ denote the covariances among the observed variables computed directly from a sample of data. To fit the model to the data, that is to estimate the free parameters in $\mathbf{A}$ and $\mathbf{P}$, we have to select parameter values that make $\mathbf{S}$ as close as possible to the model-implied covariances $\mathbf{C}$. Under the assumptions that the errors and latent variables are multinormally distributed, maximum-likelihood estimates of the free parameters in $\mathbf{A}$ and $\mathbf{P}$ is equivalent to minimizing the fitting criterion

$$\mathrm{F}(\mathbf{A,P}) = \mathrm{trace}(\mathbf{SC}^{-1}) - n + \ln(\mathrm{det}\mathbf{C}) - \ln(\mathrm{det}\mathbf{S}).$$

In the *R* package "*sem*", Fox implements a very flexible SEM, including the ability to estimate models by two-stage least squares (2SLS), and to fit general (including latent variable) models by full information maximum likelihood (FIML) assuming multinormality. At the start of simulation, we found some difficulty to use and to interpret the results obtained with this package of *R*. To explore and to understand the estimation procedure implemented in "*sem*", we tried to apply simplified models to see where wrong, so we found a small bug in the software. Following our own alert to Fox, this problem has been solved, the package has been update and we have been able to successfully use this package for our analysis.

2.4 Item Response Models (IRMs)

As we have seen in the introductory paragraph, the problem is how to obtain an objective measure of the latent construct of interest. Within the framework of Item Response Theory (IRT), several models have been proposed to synthesize data obtained from a questionnaire producing an objective measure of the latent construct. The goal of an Item Response Model (IRM) is to describe, trough a nonlinear monotonic function, the association between a respondent's underlying latent trait level and the probability of a particular item response. To produce objective measures, the available data must satisfy the following two prerequisites: unidimensionality and local independence. Unidimensionality requires that only one construct is measured, and then all the items included in the questionnaire should be closely linked to it. Local independence, for which observed items are conditionally independent of each other given an individual score on the latent variable, is obtained when the complete latent trait space is specified in the model. If the assumption of unidimensionality holds, then only a single latent trait is influencing item responses and local independence is obtained. 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 (Bond and Fox, 2001).

The first step for an IRM analysis is to define a questionnaire to capture the latent construct, paying attention that each of the aspects that identify the same construct is properly investigated by at least one item. In case of job satisfaction, for example, the questionnaire will include items relating to working conditions, relationships with colleagues or pay. The answer to each question is chosen among a set of possible categories ordered on a Likert scale; successive categories identified higher levels of latent construct, for example from dissatisfied to very satisfied. Then we numerically coded these response categories (0,1,2, ...) and, for each subject, we compute the raw score by adding together his single observed scores. The raw score cannot be treated as a measure, because of the possible responses are qualitative, not quantitative and the number assigned to response categories to produce

ordinal level data, could lead to treat them as if they were interval-level scales, without being actually. Furthermore, the presence of missing data or non-responses, heavily influence the raw score.

Originally IRMs were proposed in psychometric framework for assessing the level of individual intelligence; for this reason, in literature, the extension of latent construct is identified with the term "skills". Obviously it can be identified with other different aspects, according to context where they are applied, for example satisfaction, happiness, fear, pain, etc. Usually, these instruments are applied to conventional test responses, however they are more general and they can be applied wherever discrete data are obtained with the intention of measuring a quantitative attribute or trait.

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). Using conditional maximum likelihood (CML) estimation, both types of parameters may be estimated independently from each other. The incorporation of linear structures allows for modeling the effects of covariates and enables the analysis of repeated categorical measurements.

We now present, briefly, three general categories of IRMs for dichotomous data: One-parameter model, Two-parameter model, Three-parameter model. In Section 2.4.2 we will extend to the case of polytomous data; for further details on this class of models see Andrich (1978), Lord (1980), Wright and Masters (1982), Nering and Ostini (2010).

One-Parameter Model

In the 1-parameter logistic model, or 1-PLM, observed dichotomous item responses are a function of the ability (latent trait) possessed by the i-th subject (θ_i) and the difficulty of the j-th item (δ_j):

$$P(y_{ij} = 1) = \frac{exp[D(\theta_i - \delta_j)]}{1 + exp[D(\theta_i - \delta_j)]} = \frac{1}{1 + exp[-D(\theta_i - \delta_j)]}. \tag{2.6}$$

Latent trait scores and item difficulty parameters are estimated independently, but on the same score metric (logit scale). For dichotomous

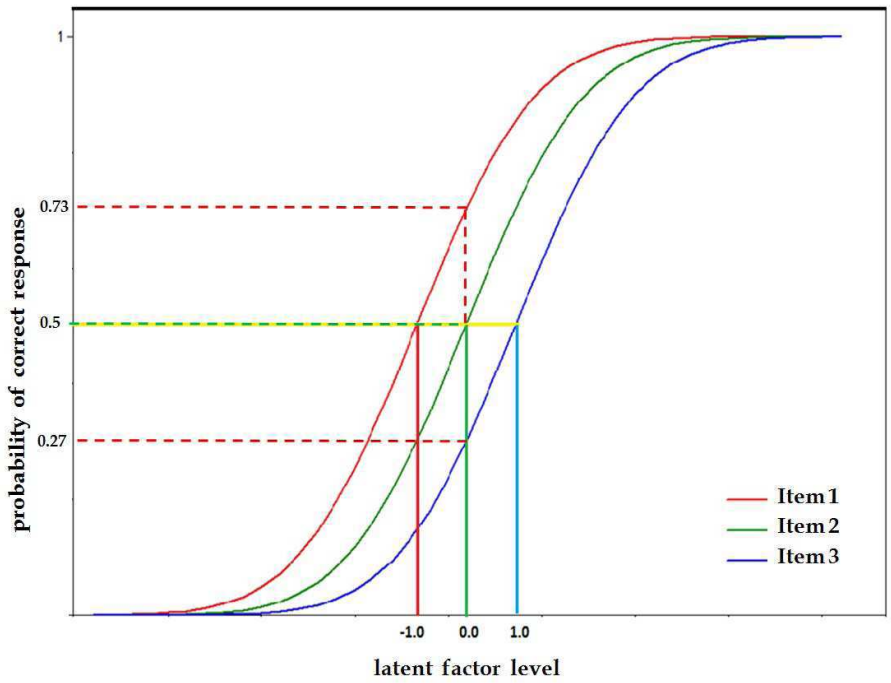


Figure 2.2. Item characteristic curves for 3 dichotomous items

data, the item difficulty parameter corresponds to the latent trait level at which a subject has a 50% probability of success; in general, if a person has a latent trait level greater then an item difficulty, his probability of correctly responding to this item is greater than 50%, otherwise it is less than 50%. The scaling factor D, if it is set to 1.7, is used to make the logistic function essentially the same as the normal ogive model.

We can plot $P(y_{ij} = 1)$ as a function of ability, obtaining a smooth S-shaped curve known as Item Characteristic Curve (ICC): the probability of correct response is near zero at the lowest levels of ability, it increases until 1 at the highest levels of ability. Figure 2.2 presents the ICCs for 3 dichotomous items, with three different item difficulties (-1, 0, 1). Because of in the 1-PLM only the item difficulty can vary, the s-shaped curves have the same slope so they are parallel. The probability of a correct response to the easiest item ($\delta_1 = -1$) for someone of average ability ($\theta = 0$) is 0.73, whereas the probability for the second item ($\delta_2 = 0$) is 0.50 (this is true by definition given that $\theta = \delta$), and the probability for the hardest items ($\delta_1 = 1$) is 0.27.

The most popular 1-PLM is the *Rasch model*, that we will present in the Section 2.4.1. In the Rasch model, the log odds of a person endorsing in the higher category is simply a function of difference between the subject latent trait level and the item difficulty. A nice feature of the Rasch model is that observed raw scores are sufficient for estimating latent trait scores using a nonlinear transformation; in other IRMs, the raw score is not a sufficient statistic.

Two-Parameter Model

The 2-parameter logistic model (2-PLM) is an extensions of the 1-PLM, in which we estimate, besides the item difficulty parameters δ_j, the item discrimination parameters α_j.

$$P(y_{ij} = 1) = \frac{exp[D\alpha_j(\theta_i - \delta_j)]}{1 + exp[D\alpha_j(\theta_i - \delta_j)]} = \frac{1}{1 + exp[-D\alpha_j(\theta_i - \delta_j)]}. \qquad (2.7)$$

The discrimination parameter, typically ranges from 0.5 to 2, is similar to an item-total correlation: high values of α_j indicate items that are better able to discriminate between contiguous trait levels near the inflection point, their ICCs have a steeper slope.

If we employ the normal distribution, we obtain the two-parameter normal ogive model (Lord, 1952):

$$P(y_{ij} = 1) = \phi[\alpha_j(\theta_i - \delta_j)] = \int_{-\infty}^{\alpha_j(\theta_i - \delta_j)} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz,$$

where ϕ is the standard normal cumulative distribution function.

An important characteristic of the 2-PLM is that the difference between the subject's trait level and the item difficulty $(\theta_i - \delta_j)$ has a greater impact on the probability of correct answer for highly discriminating items than for less discriminating items.

The maximum likelihood estimation procedure is based on finding the item characteristic curve that best fits the observed proportions of correct response. In Ferrando and Chico (2007) it is possible to fine more details on this estimation method.

Three-Parameter Model

Sometimes, we could be interested in adjusting the estimations for the impact of chance on observed scores; this because, for example, in ability testing it is possible to get an answer right by chance, raising the lower asymptote of the function. To solve this problem, in a 3-parameter logistic model, or 3-PLM, we add a pseudo-guessing parameter (c), and the probability of a correct response of the i-th subject to the j-th dichotomous item is

$$P(y_{ij} = 1) = c + (1 - c) \times \frac{exp[D\alpha_j(\theta_i - \delta_j)]}{1 + exp[D\alpha_j(\theta_i - \delta_j)]}.$$

When $\theta_i = \delta_j$, the response probability is $\frac{1+c}{2}$. The parameter estimation is similar to 2-PLM, to deepen this argument see Baker (1985) and Baker and Kim (2004).

2.4.1 The Rasch Model (RM)

This model is named after Georg Rasch, a famous Danish mathematician and statistician early last century lived. The RM has been conceived for dichotomous data, even if many extensions have been studied for polytomous data (some of that will be presented in Section 2.4.2).

When Rasch implemented its model, he was concerned principally with the measurement of individuals, rather than with distributions among populations. He want to obtain a measurement congruent with the physical measurement characteristics (for example invariant comparison), without making any assumptions about the distribution of trait levels in a population. He summarized the principle of invariant comparison as follows (Rasch, 1961):

> *The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; and it should also be independent of which other stimuli within the considered class were or might also have been compared. Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the*

> *class considered were instrumental for the comparison; and it should also be independent of which other individuals were also compared, on the same or some other occasion.*

The RM meets this principle as it has been implemented in order to obtain a separation between item and person parameters, so when we estimate item parameters, we can ignore person parameters, and viceversa. To achieve this, it is necessary to use a conditional maximum likelihood estimation, in which the response space is partitioned according to person total scores. As we have already said, in these models the item (person) raw score is the sufficient statistic for the item (person) parameter: the person total score contains all information available within the specified context about the individual, and the item total score contains all information with respect to item, with regard to the relevant latent trait.

In the RM, the probability of a correct response is modeled as a logistic function of the difference between the person and item parameter:

$$P(y_{ij} = 1) = \frac{exp(\theta_i - \delta_j)}{1 + exp(\theta_i - \delta_j)}; \tag{2.8}$$

the higher is a person's ability relative to the difficulty of an item, the higher is the probability, for this person, to correctly response on that item. In the original Rasch formulation, we don't have the constant D as in (2.6), that has been inserted in the 1-PLM to make comparable logistic function with the normal ogive model.

An important step of these models is the item calibration, during which item locations are scaled: for example, items with the smallest proportion of correct responses are the most difficult and hence they have the highest items locations δ_j. Once item locations are determined, we can the measure person locations θ_i on that scale; so person and item locations are estimated on a single scale (logit).

2.4.2 The Polytomous Item Response Model (PIRM)

So far we have developed suitable models for dichotomous items, but in fact most of the questionnaires provide a choice of polytomous responses (with c_j, $j = 1, 2, \ldots, p$, possible answers). Many authors studied this problem; we will briefly presented the essential features of three of the most popular proposed models: the Graded Response Model (GRM), the Partial Credit Model (PCM), and in more detail the Rating Scale Model (RSM), as this last model is used in the simulation part of the present work.

Graded Response Model (GRM)

The graded response model, originally introduced by Samejima (1969), is an extension of the 2-PL (2.7) for items with different numbers of response categories. It assumes a normal cumulative distribution function for the item response function, so the probability that the i-th subject meets maximum class k (i.e he/she can response a categories $\leq k$) is

$$P(y_{ij} \leq k) = \frac{1}{\sqrt{2\pi}} \int_{-\alpha_j(\theta_i - \delta_{jk})}^{\infty} e^{\frac{-z^2}{2}} dz = \phi_{jk}(\theta_i).$$

The item parameter δ_{jk} includes the item slope parameter and the $(c_j - 1)$ between category threshold parameters τ_{jk}, representing the latent trait level necessary to respond above threshold with 50% probability. For GRM, one operating characteristic curve needs to be estimated for each between category threshold, they represent the probability of responding in a particular category conditional on trait level.

Partial Credit Model (PCM)

The partial credit model is an extension of the RM in (2.8) to polytomous items, thus item slopes are assumed to be equal across items. As in GRM, items can have different numbers of categories. The person probability of responding in category k is a function of the difference between his trait level

and a category intersection parameter (Masters, 1982):

$$P(y_{ij} = k) = \frac{exp\left[k(\theta_i - \delta_j) - \sum_{l=0}^{k} \tau_{jl}\right]}{\sum_{g=0}^{c_j-1} exp\left[g(\theta_i - \delta_j) - \sum_{l=0}^{g} \tau_{jl}\right]}, \qquad k = 0, 1, \ldots, c_j. \tag{2.9}$$

The intersection parameters, τ_{jl}, represent the trait level at which a response the l-th category becomes more likely than a response in the previous one.

2.4.3 The Rating Scale Model (RSM)

The Rating Scale Model (RSM) was derived by Andrich (1978) from the idea of Rasch (1961), focusing on the use of Likert scales in psychometrics. The PCM has an identical mathematical structure, but was derived from a different starting point at a later time. The two main features of this model are that items have the same number of categories (contrary to PCM) and the difference between any given threshold location and the mean of the threshold locations is equal or uniform across items.

The RSM is a probabilistic measurement model in which the raw scores are sufficient statistics for the parameters of the models (a key features of the Rasch models); moreover it permits, by empirical tests, to verify the hypothesis that the response categories represent increasing levels of latent construct, so that they are correctly ordered. In 2005, Andrich has significantly developed the problem of disordered threshold locations, proposing different solutions depending on the specific context.

In this model, the location parameter δ_j represents the average difficulty for a particular item relative to the category intersections; each item is assumed to provide the same amount of information and have the same slope. Encoding the c response categories ordered as $0, 1, 2, \ldots, c-1$, (all the items have the same number of categories), the RSM probability for the

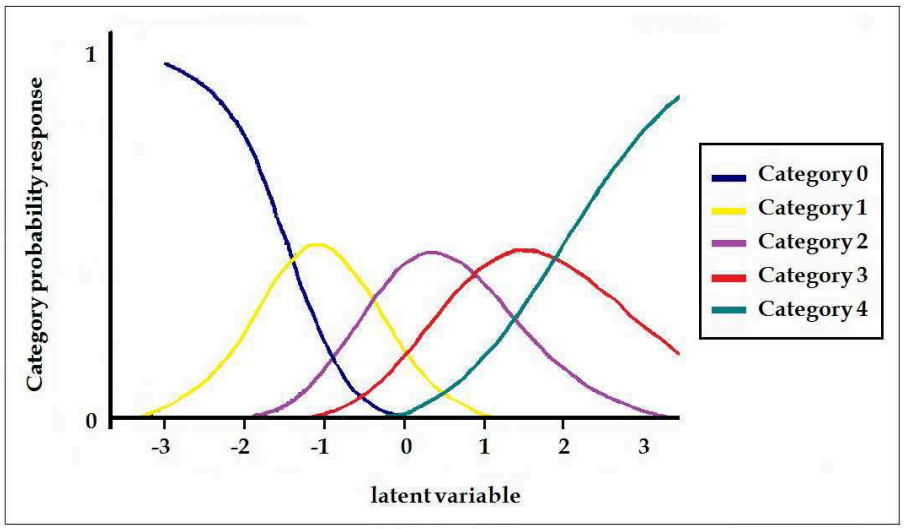


Figure 2.3. Category Probability Curves for an item with 5 ordered categories

i-th subject to answer at the j-th item the k-categories is similar to (2.9)

$$P(y_{ij} = k) = \frac{exp\left[k(\theta_i - \delta_j) - \sum_{l=0}^{k} \tau_l\right]}{\sum_{g=0}^{c-1} exp\left[g(\theta_i - \delta_j) - \sum_{l=0}^{g} \tau_l\right]}, \qquad k = 0, \ldots, c,$$

where θ_i identifies the level of latent aspect possessed by the subject i, δ_j is the level of difficulty of the j-th item, while the threshold parameter τ_l (with $\tau_0 \equiv 0$ and $\sum_{l=0}^{c} \tau_l = 0$) quantifies the difficulty of choosing the l-th answer rather than the previous one.

The probability of a given category as a function of person location is referred to as a Category Probability Curve (CPC); an example for an item with five categories, scored from 1 to 5, is shown in Figure 2.3.

The individual latent aspect θ_i, the item difficulty δ_j and thresholds τ_l are measured on a logit scale (log odds unit), and so we can then sort them on the basis of measurements from both subjects (from the most satisfied to least satisfied) and 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.

The item score is the count of the number of threshold locations on the latent trait surpassed by the individual. We do not make this measurement counts in a literal sense; rather, threshold locations on a latent continuum are usually inferred from a matrix of response data through an estimation process such as Conditional Maximum likelihood estimation.

Extended Rasch Models

For the simulation study, we have adopted a slightly different formulation of the RSM, belonging to the family of Extended Rasch Models (ERM), proposed by Mair and Hatzinger (2007a) and implemented in the *R* package "eRm". ERMs are a very general framework which includes several models, but the formalization, for the case under study, is almost equal to that shown in (2.4.3):

$$P(y_{ij} = k) = \frac{exp[k(\theta_i + \delta_j) + \omega_k]}{\sum_{l=0}^{c-1} exp[l(\theta_i + \delta_j) + \omega_l]}, \tag{2.10}$$

where ω_k are the category parameters ($\omega_k = \sum_{l=0}^{k} \tau_l$).

In this model, the authors use a conditional maximum likelihood (CML) approach, which permits to estimate the item parameters δ_j without estimating the person parameters θ_i by conditioning the likelihood on the sufficient person raw score. Thanks to a design matrix approach which allows to the user to impose repeated measurement designs as well as group contrasts, the item parameter may differ over time with respect to certain subgroups.

The likelihood expressions

In literature have been proposed several estimation methods for IRM, the most used are conditional maximum likelihood (CML) and marginal maximum likelihood (MML) estimation, that requires to specify a density function, $f(\theta)$, for the person parameters; CML and MML are asymptotically equivalent and provide consistent estimators (Pfanzagl,

1994; Hatzinger, 2008).

For the ERM, the authors use CML approach for the desirable properties of the estimators and because it stays close to the concept of specific objectivity (Rasch 1960; Fisher Jr. 1992). CML estimation assumes that the person's raw score $r_i = \sum_{j=1}^{p} y_{ij}$ is a sufficient statistic. Thus, considering the conditional likelihood with respect to

$$r = (r_1, r_2, \ldots, r_n) \tag{2.11}$$

the person parameters θ disappear from the likelihood equation, thus, leading to consistently estimated item parameters $\hat{\delta}$. To ensure identifiability, we have to impose some restrictions, for example in the RM one item parameter has to be fixed to 0 and it is considered as baseline difficulty. In addition, in the RSM the category parameter τ_0 is constrained to 0. The conditional log-likelihood equation for the RSM is

$$\log L_c = \sum_{j=1}^{p} \sum_{k=1}^{c_j} y_{ijk}\theta_j - \sum_{r=1}^{r_{max}} n_r \log \gamma_r,$$

where r_{max} is the maximal raw score, n_r is the number of subjects with the same raw score, the γ-terms are the elementary symmetric functions and y_{ijk} is 1 if $y_{ij} = k$, 0 otherwise. An elaborated derivation of these terms for the ordinary RM can be found in Fischer (1974) and an overview of various computation algorithms is given in Liou (1994). For all steps, the first and second order derivatives and the solution algorithm, see Mair and Hatzinger (2007a and 2007b).

2.4.4 The Analysis of Reliability

When we use an IRM, it is important to evaluate the appropriateness of the found factor structure with an evaluation of the reliability of the obtained measures (Tarkkonen and Vehkalahti, 2005). There is an important difference between validity and reliability. To better express it, we refer to the definition of these concepts given by Vogt in his Dictionary of Statistics & Methodology

(1993).

- *Reliability*: Freedom from measurement error. In practice, this boils down to the consistency or stability of a measure or test from one use to the next.

- *Validity*: A term to describe a measurement instrument or test that accurately measures what it is supposed to measure; the extent to which a measure is free of systematic error. Validity requires reliability, but the reverse is not true.

Gulliksen (1945) and Loevinger (1954) studied the attenuation paradox, for which the validity is reduced by increasing the reliability, but it is important to have measures with a high reliability and we can do it increasing the inter-item correlation (IIC). The problem is that when we increase too much the IIC, all items are so higly correlated that effectively they reduce to one item, so we reduce the validity (Kuhn, 1970).

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

$$\text{Reliability} = \frac{\text{True variance}}{\text{Observed variance}}$$

If measures and errors are uncorrelated, then

$$\text{Observed variance} = \text{True variance} + \text{Error variance}$$

Cronbach's alpha, 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 of the other indices (for example the Kuder-Richardson, KR, coefficients):

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

where p is the number of items. σ_t^2 is the variance of r_i, the observed raw scores for the current sample of persons (2.11), σ_j^2 is the variance of the j-th item for the current sample of persons, $\bar{r}_j$ is the means of the answers at the j-th item for the current sample of persons:

$$\begin{aligned} \sigma_j^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_{ij}-\bar{r}_j)^2 \qquad \sigma_t^2 = \frac{1}{n}\sum_{i=1}^{n}(r_i-\bar{r})^2 \\ \bar{r}_j &= \frac{1}{n}\sum_{i=1}^{n} y_{ij} \qquad\qquad\qquad \bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i \end{aligned}$$

Cronbach's alpha describes the internal consistency of groupings of items; an high value of the index indicates that the responders express a coherent position on each item belonging to the same dimension. In literature, we find that a barely acceptable level of internal consistency and appropriateness of test is $\alpha \geq 0.7$ (Nunnally and Bernstein, 1994).

Checking the internal consistency of each test subset, we can deepen the study of the structure factors and define the construct validity of the scale. Novick and Lewis (1967) showed that Cronbach's alpha is a lower bound to the true reliability, so we can consider Cronbach's alpha as a conservative estimate of the reliability.

Often in Rasch Measurement the person separation index is used instead of reliability indices. **Rasch Person Reliability index (RPRI)** indicates the replicability of person ordering that we could expect if at this sample of persons were given another parallel set of items measuring the same construct. This index is a ratio of the latent construct variance (SA^2) and the measure variance (SD^2):

$$\text{RPRI} = \frac{SA^2}{SD^2} = 1 - \frac{MSME}{SD^2}, \tag{2.12}$$

where the Mean Square Measurement Error (MSME) is

$$MSME = \frac{1}{n}\sum_{j=1}^{n} S_j^2, \tag{2.13}$$

and $SA^2 = SD^2 - MSME$. The error variance of all measures obtained with IRMs proposed in the previous paragraphs 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. RMs provide a direct estimate (SE_n) of the model error variance for each estimate of a subject's ability and an item's difficulty (Wright and Masters, 1982; Wright, 1999), that gives a quantification of the precision of every person measure and item difficulty. The level of measurement error is not uniform across the range of a test, but is generally larger for more extreme scores. Person reliability requires not only ability estimates well targeted by a suitable pool of items, but also a large enough spread of ability across the sample so that the measures demonstrate a hierarchy of ability (person separation) on this construct. Unfortunately, the correlation-based reliability index (category at which RPRI belongs to) are nonlinear, so the improvement form 0.7 to 0.8 is not twice the improvement from 0.9 to 0.95.

Conventionally, only a Reliability index is reported. The relationship between raw-score-based reliability index (Cronbach alpha) and measure-based reliability index (RPRI) is complex (Linacre, 1997; Schumacker and Smith, 2007); in general, Cronbach alpha overestimates reliability, RPRI underestimates it. The big differences between Score and Measure reliability occur when there are:

- extreme scores, that increase score reliability, but decrease measure reliability;
- missing data, that always decreases measure reliability. If the missing data are imputed at their expected values (in order to make conventional reliability formulas computable), they increase score reliability.

Chapter 3

The Estimation Procedures: A Simulation Study

3.1 Introduction

When we analyze socio-economic phenomena, we often have observed realizations of the latent aspects, which are linked by different relations of dependence that we are very interested in finding out. As we have said in chapter 2, the study of all the presented models originates from this concrete problem: to measure subjective perceptions and to understand their dependencies. The objective of the researches is, therefore, to estimate the parameters of a regression model for latent variables, that are measured indirectly with sets of Likert-type items (ordinal scale). In the second chapter, we have seen the two different approaches for the analysis of categorical observed data: the Underlying Variable Approach (Section 2.2.1) and the Item Response Models (Section 2.2.2). Each of these approaches is linked to a different estimation procedure: we will present the One-step Procedure for UVA, and the Two-step approach for IRMs in Section 3.2. The purpose of the simulation study is to compare, on the same data, the results obtained using these different estimation procedures.

In Section 3.3, we will present the simulation design, entering some details about the data generation. Finally in Section 3.4, we will report the results

of the simulation study and we will make some conclusion considerations.

3.2 The Estimation Procedures

The starting point is the analysis of a complex phenomena (at least two latent factors, but often we have many more) using a series of data collected from a questionnaire with Likert-type items. Many authors are interested in studying the relations between latent constructs, between the most famous we remember Goldberger and Duncan (1973), Skrondal and Rabe-Hesketh (2004), Bollen and Curran (2006), and in a multilevel optical Hox (2002).

For example, we can think of a study addressing the workers of social cooperatives (Carpita, 2009): the data are collected trough a questionnaire designed to investigate, through numerous questions, 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 equity (distributing resources, balance between what the worker gives the organization and what that it receives). Besides getting a good measure of these constructs, the researcher is obviously interested in understanding the relationships between them.

Obtain "good" measures (see Section 2 and 2.4.4) and to assess the dependence relationships between the constructs they represent, are the two main research interests for the analysis of latent variable models with psychological traits. These two objectives (measures and dependence links) may be combined into a single estimation procedure or developed sequentially one at a time.

- If we combine the two interests in a single model, we have a ***One-step procedure***. This procedure requires the researcher to specify 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 a ***Two-step procedure*** when we estimate the measures and their dependence in two different phases. In the first step, we estimate, separately, 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 dependance relationships. It, however, does not allow a thorough analysis of obtained measures in IRT perspective, the strength of Two-step procedure.

The properties of the Two-step procedure have not yet been adequately detailed in the literature and we did not find significant contributions in this regard. One of the goal of this research is 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 methods. Crucial element is to find a correct model that considers that the measures obtained by the first step are affected by measurement errors.

We have implemented an articulated simulation study (Section 3.3) 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 also evaluating that it allows better control in both phases: measures construction and regression model.

Moreover, in the Two-step method we use a reliability index to estimate the variance of measurement error, not yet widespread in the literature: the Rasch Person Reliability Index (2.12). This index can conceptually be an element of connection between the first and second stage of Two-step procedure, as it is calculated together with the measures in the first step and it determines the magnitude of measurement error in the second one.

3.2.1 The One-step Procedure

This procedure is combined with the UVA approach, showed in Section 2.2. 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 a structural equation model. We will refer to the Muthén SEMs, implemented in M*plus* (Muthén and Muthén, 2007).

We have seen that two main components of models are distinguished in SEM: the Structural Model, showing potential causal dependencies between endogenous and exogenous variables, and the measurement model showing the relations between latent variables and their indicators. In the simple case of two latent factors with three and two categorical indicators, respectively, we have:

The **Structural Model**

$$\begin{aligned} \theta_1 &= \zeta_1 \\ \theta_2 &= b\theta_1 + \zeta_2, \end{aligned} \tag{3.1}$$

The **Measurement Model**

$$\begin{aligned} Y_1^* &= \lambda_{11}\theta_1 + \epsilon_1 \\ Y_2^* &= \lambda_{12}\theta_1 + \epsilon_2 \\ Y_3^* &= \lambda_{13}\theta_1 + \epsilon_3 \\ Y_4^* &= \lambda_{24}\theta_2 + \epsilon_4 \\ Y_5^* &= \lambda_{25}\theta_2 + \epsilon_5, \end{aligned}$$

where $\epsilon_j \sim N(0,1)$ and λ_{ij} are the loadings relating the indicators to their

constructs. For each ordinal indicator item, we have

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

It is important to underly that all the model error terms, ζ_h and ϵ_j, 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 .

We implemented two different estimation methods:

- Structural Equation Model standard (**SEMstd**),
- Structural Equation Model with IRT approach (**SEMirt**).

In Appendix we have included the scripts of M*plus* used to apply these methods.

SEMstd is the simple Structural Equation Model just seen in equations (3.1) and (3.2). We are interested in the estimation of the coefficient regression b and of the threshold parameters $\tau_{h,k}$. We have to remember that, in RSM, the threshold parameters are equal for all the items referring to the same latent construct, so $\tau_{h,k}$ is the threshold parameter for the k-th categories for all the items referring to the h-th construct. To homogenize the comparisons with the results obtained through other estimation methods and with different experimental conditions, we always standardize all the estimated variables and coefficients, so instead of talking about regression coefficient b, we will refer to beta weights β, presented in the first Appendix (A.1).

SEMirt is a version of previous the model, inspired by the work of Gibbons et al. (2007), that introduces the structure of IRMs in SEM. M*plus* do 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, formally, are 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 IRMs, we have to impose that all relevant loadings be equal to 1. So, we have:

The **Structural Model**

$$\begin{aligned}
\theta_1 &= \zeta_1 \\
\theta_2 &= \beta\theta_1 + \zeta_2 \\
\Delta_1 &= \delta_1 + \zeta_3 \\
\Delta_2 &= \delta_2 + \zeta_4 \\
\Delta_3 &= \delta_3 + \zeta_5 \\
\Delta_7 &= \delta_7 + \zeta_6 \\
\Delta_5 &= \delta_5 + \zeta_7,
\end{aligned}$$

with $\text{var}(\Delta_j) \equiv 0$. The parameter δ_j, the mean of the variable Δ_j, represents the difficulty of the j-th item.

The **Measurement Model**

$$\begin{aligned}
Y_1^* &= \theta_1 + \Delta_1 + \epsilon_1 \\
Y_2^* &= \theta_1 + \Delta_2 + \epsilon_2 \\
Y_3^* &= \theta_1 + \Delta_3 + \epsilon_3 \\
Y_4^* &= \theta_2 + \Delta_4 + \epsilon_4 \\
Y_5^* &= \theta_2 + \Delta_5 + \epsilon_5,
\end{aligned}$$

where $\epsilon_j \sim N(0, 1)$.

With this model, we can estimate the standardized regression coefficient β, the threshold parameters $\tau_{h,k}$ and, in addition to the previous model, the item difficulty parameters δ_j (through the means of Δ_j).

For SEMs, several indices of goodness of fit have been proposed, but it is

not possible to proceed to the analysis of reliability and make 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.

3.2.2 The Two-step Procedure

This procedure is combined with the Item Response Theory approach (see Section 2.4), which focuses on observed variables. The first step of this method consists in estimating, 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.

In the first step, we divide the items in many subdatabases as there are latent factors, so that items belonging to each subdataset refer to a single construct. This is necessary to respect the character of unidimensionality required by IRMs. Once the various measures are constructed, for each of them we can analyze the goodness of results. All IRMs include an analysis phase where researchers have to determine if the constructed measure mets all the main features of these models (for example unidimensionality and category proper order), and to estimate their reliability.

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

Because of the properties of the Two-step procedure have not yet been adequately detailed in the literature, we define two different method for this procedure, with the intention of being able to evaluate the essential characteristics of the estimators on simulated data:

- Rating Scale Model - Linear Regression Model with Measurement Error

(**RSM-LRMme**);

- Rating Scale Model - Standard Linear Regression Model (**RSM-LRM**).

The first method has been implemented searching 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 methods. Crucial element is to find a correct model that considers that the measures obtained by the first step are affected by measurement errors. Instead the second method, more traditional, aware that the measures are affected by measurement error.

In Appendix we have included the scripts of R used to apply these methods.

The first step of both methods is the same: for each of the m constructs, we estimate its measure, Y^*, through an extended RSM (2.10), and their person reliability, using the standard errors of the person parameters (Mair and Hatzinger, 2007b). In this phase, we use the R package eRm of Mair and Hatzinger (2007a).

Before moving to the next step, we standardize all the estimated measures and the other quantities involved in the model. In this way, the regression coefficients that we will estimate in the second step will be the beta weights, β, presented in the first Appendix (A.1), and they will be comparable with that 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.

In **RSM-LRMme**, we implemented a linear regression model taking into account, to estimate the regression coefficients, 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, that we have amply described in the first chapter (see Section 1.3.1 at page 7). The regression

equation is:

$$\theta_2 = \beta\theta_1 + \zeta_2,$$

where

$$\begin{aligned} Y_1^* &= \lambda_{11}\theta_1 + \epsilon_1 \\ Y_2^* &= \lambda_{22}\theta_2 + \epsilon_2, \end{aligned}$$

with the loadings λ_{ij}.

To estimate this model, we used the RAM (in Section 2.3.5), implemented in the *R* package "*sem*" of Fox (2006). As measurement error variance estimates, we use the variance error estimates derived in the previous step (2.13). We decide to refer to the Rasch Person Reliability Index (2.12), still not used in scientific papers to estimate the variance of measurement error, because it can conceptually be an element of connection between the first and second stage of Two-step procedure, as 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:

$$Y_2^* = \beta Y_1^* + \varepsilon_1.$$

What we expect, and we will verify with the simulation study, is that estimated regression coefficients $\hat{\beta}$, obtained with the latter method, are lower than those obtained with the **RSM-LRMme** method, because of the attenuation effect of measurement error (1.7).

3.3 The Simulation Design

The objective of this study is to compare the results obtained applying the 4 different analytical methods, previously described, to the same data. We created many different simulated databases in order to evaluate the obtained

estimates, knowing the real value of the parameter of interest. We indica with p be the total number of the items, p_1 is the number of the indicators for the first construct, p_2 that for the second one and p_3 that for the third one.

We imagined two scenarios: the first is simple with only two latent factors and a dependence relationship of the second versus the first; in the second one we have considered three latent factors, where one is dependent from the other two.

For the first scenario, the parameter p_2 is always fixed at 5, while $p_1 = 5$ for case A, and $p_1 = 10$ for case B. The model generating the data is therefore (Figure 3.1):

The **Structural Model**

$$\begin{aligned} \theta_1 &= \zeta_1 \\ \theta_2 &= \beta_{12}\theta_1 + \zeta_2 \\ \Delta_1 &= \delta_1 + \zeta_3 \\ \vdots &= \vdots \\ \Delta_p &= \delta_p + \zeta_p + 2, \end{aligned}$$

with $\text{var}(\Delta_j) \equiv 0$, $\zeta_1 \sim N(0,1)$ and ψ_2 fixed such that $\theta_2 \sim N(0,1)$.

The **Measurement Model**

$$\begin{aligned} Y_1^* &= \theta_1 + \Delta_1 + \epsilon_1 \\ \vdots &= \vdots \\ Y_{p_1}^* &= \theta_1 + \Delta_{p_1} + \epsilon_{p_1} \\ Y_{p_1+1}^* &= \theta_2 + \Delta_{p_1+1} + \epsilon_{p_1+1} \\ \vdots &= \vdots \\ Y_p^* &= \theta_2 + \Delta_p + \epsilon_p \end{aligned}$$

with $\epsilon_j \sim N(0,1)$.

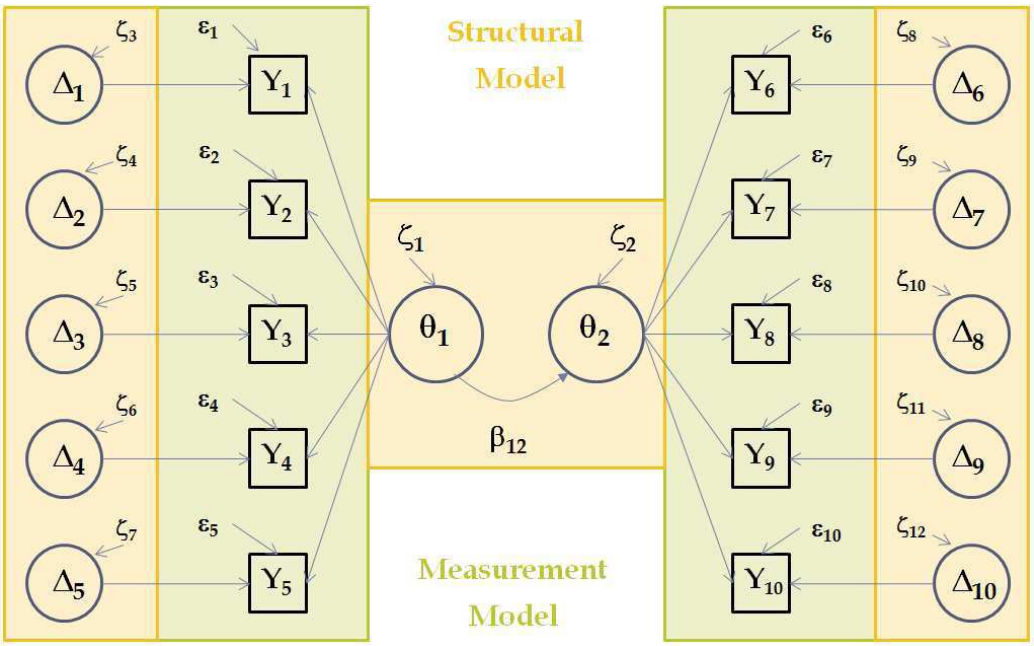


Figure 3.1. **SEMirt** 2 latent factors M*plus* model

For the second scenario, the parameters p_2 and p_3 are always equal to 5, while p_1 is equal to 5 for case An and to 10 for case B. The model generating the data is (Figure 3.2):

The **Structural Model**

$$\begin{aligned}
\theta_1 &= \zeta_1 \\
\theta_2 &= \beta_{12}\theta_1 + \beta_{32}\theta_3 + \zeta_2, \\
\theta_3 &= \zeta_3 \\
\Delta_1 &= \delta_1 + \zeta_4 \\
\vdots &= \vdots \\
\Delta_p &= \delta_p + \zeta_{p+3},
\end{aligned}$$

with $\text{var}(\Delta_j) \equiv 0$. Both ζ_1 and ζ_1 are distributed as a standard normal. ψ_2, that is the variance of ζ_2, is fixed so that the variance of θ_2 is equal to 1 and we have $\theta_2 \sim N(0, 1)$. ψ_2 changes according to the value of the regression coefficients β_{12} and β_{32}.

The **Measurement Model**

$$\begin{aligned}
Y_1^* &= \theta_1 + \Delta_1 + \epsilon_1 \\
\vdots &= \vdots \\
Y_{p_1}^* &= \theta_1 + \Delta_{p_1} + \epsilon_{p_1} \\
Y_{p_1+1}^* &= \theta_2 + \Delta_{p_1+1} + \epsilon_{p_1+1} \\
\vdots &= \vdots \\
Y_{p_1+5}^* &= \theta_2 + \Delta_{p_1+5} + \epsilon_{p_1+5} \\
Y_{p_1+6}^* &= \theta_3 + \Delta_{p_1+6} + \epsilon_{p_1+6} \\
\vdots &= \vdots \\
Y_p^* &= \theta_3 + \Delta_p + \epsilon_p
\end{aligned}$$

with $\epsilon_j \sim N(0, 1)$.

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 the number of indicators for the first regressor: p_1 is equal to 5, in the case A, and equal to 10, in the case B. The parameter p_2 (and p_3 for the second scenario) is fixed at 5. 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. Each indicator is a categorical variable with 5 ordered categories.

2. We impose ψ_2, the structural error variance of the dependent latent variable θ_2, to be equal to 0.1, 0.3, 0.5, 0.7, or 0.9. In this way, we can define the strength of the dependence link between the latent variables: having fixed the variance of the latent dependent variable θ_2 equal to 1, when we increase ψ_2, we reduce the dependence relationship between of θ_3 from θ_2 and θ_1.

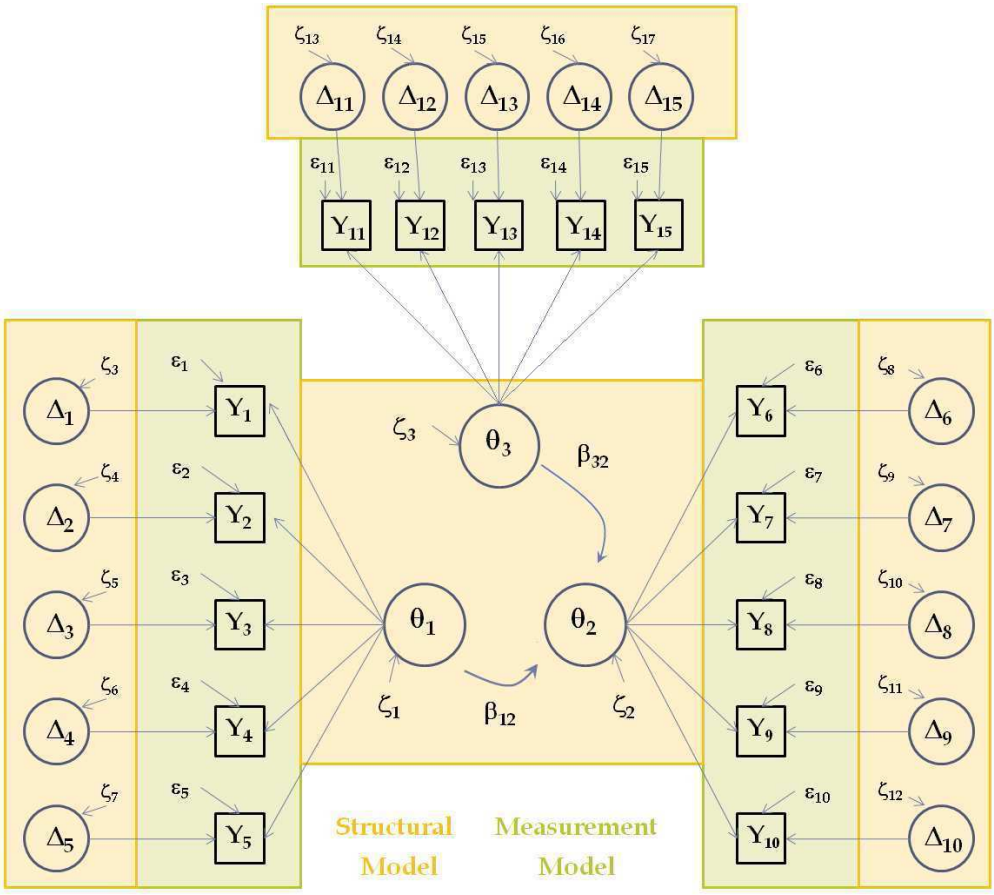


Figure 3.2. **SEMirt** 3 latent factors M*plus* model

3. Only for the second scenario, we changed the intensity of dependence of θ_2 from θ_1 and θ_3. We have

$$\theta_2 = \beta_{12}\theta_1 + \beta_{32}\theta_3 + \zeta_2.$$

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

$$\begin{aligned} \text{var}[\theta_2] &= \beta_{12}^2 \text{var}[\theta_1] + \beta_{32}^2 \text{var}[\theta_3] + \text{var}[\zeta_2] \\ &= \beta_{12}^2 \psi_1 + \beta_{32}^2 \psi_3 + \psi_2. \end{aligned}$$

$\text{var}[\theta_2]$ is fixed at 1, we set ψ_2 equal at one of five levels indicated at point 1, so $\kappa = (1 - \psi_2)$ represents the total dependence link of θ_2 from the regressor constructs. We create 3 different situations:

- Case $(\frac{1}{4}; \frac{3}{4})$, where $\beta_{12} = \sqrt{\frac{\kappa}{10}}$ and $\beta_{32} = 3\beta_{12}$. In this case the dependence of θ_2 from θ_1 is three times more intense than that from θ_3.
- Case $(\frac{1}{2}; \frac{1}{2})$, where $\beta_{12} = \beta_{32} = \sqrt{\frac{\kappa}{2}}$, so the dependence of θ_2 from θ_1 has the same intensity than that from θ_3.
- Case $(\frac{3}{4}; \frac{1}{4})$, where $\beta_{12} = 3\sqrt{\frac{\kappa}{10}}$ and $\beta_{32} = \sqrt{\frac{\kappa}{10}}$. We create this case only when $p_1 = 10$.

Obviously, when we change these three groups of parameters, we change also the values of the regression coefficients, β_{12} and β_{32}. For example, in the second scenario, with $\psi_2 = 0.1$, in the case $(\frac{3}{4}; \frac{1}{4})$ we have $\beta_{12} = 0.9$ and $\beta_{32} = 0.3$; while in the case $(\frac{1}{2}; \frac{1}{2})$ we have $\beta_{12} = \beta_{32} = 0.67082$.

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 N and n. Because, as we saw in Chapter 2 (page 29), SEMs are large samples methods, we fixed the sample size to 1000.

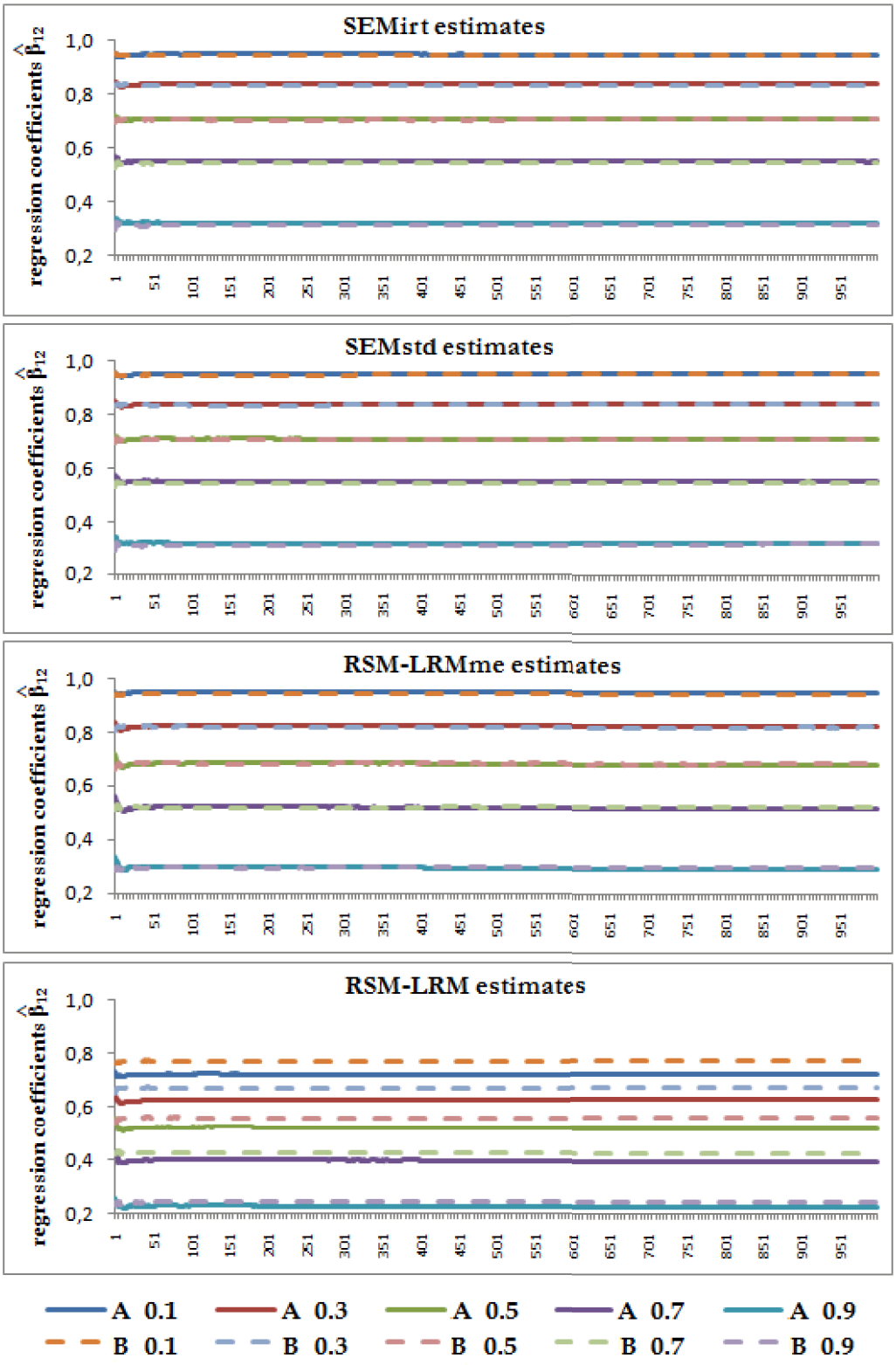


Figure 3.3. Stability of the simulation results: mean estimates of regression coefficient β_{12} in the first scenario for all 10 parameter configurations

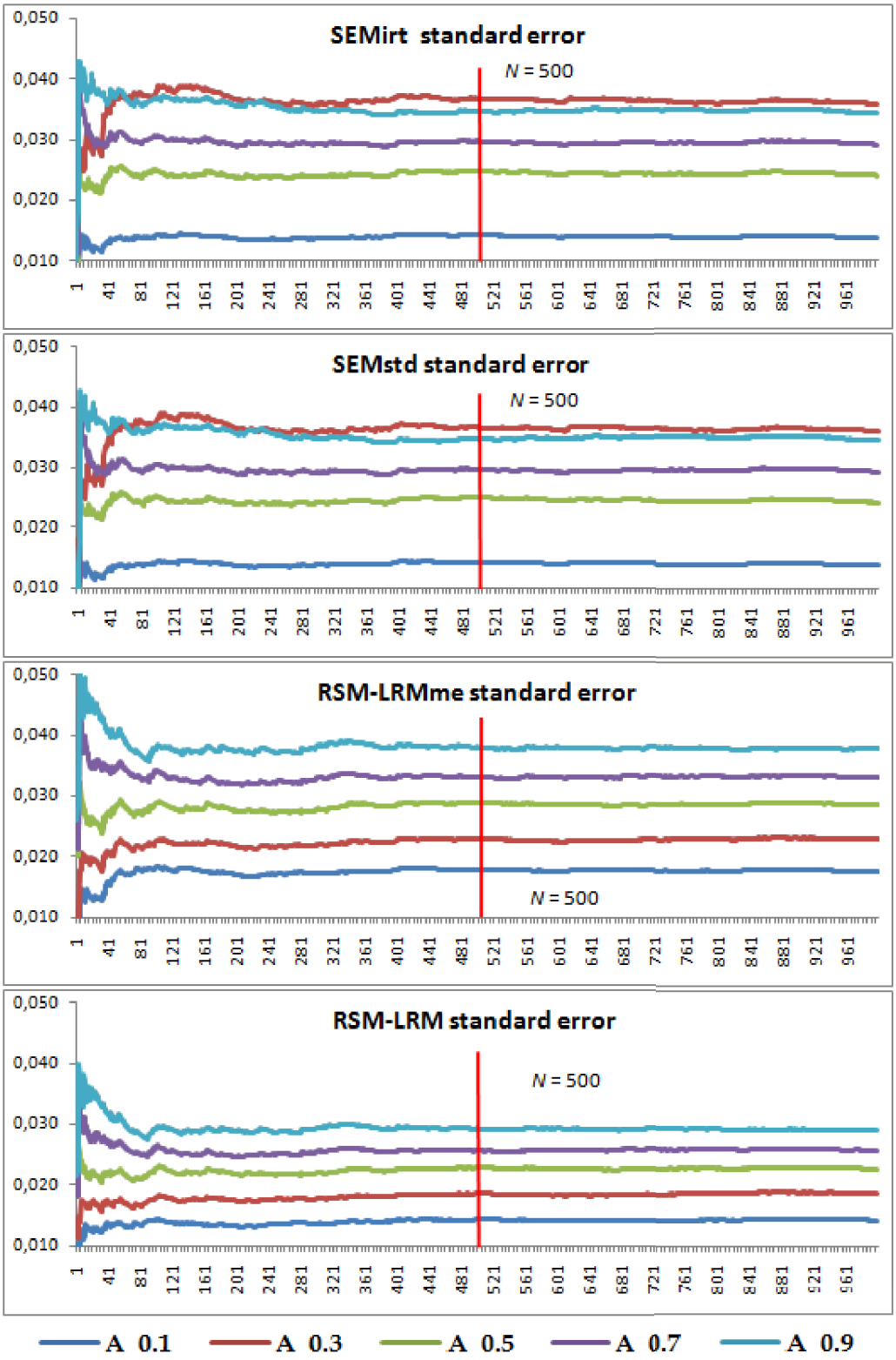


Figure 3.4. Stability of the simulation results: mean estimates of standard errors of regression coefficient β_{12} in the first scenario for all 5 parameter configurations of the case A

To decide the number of samples, we tested the stability of estimates for each of the 4 different methods above described. To simplicity, we report only the case of the first scenario (with two constructs). In Figure 3.3, we see the simulated mean estimates of the regression coefficient β_{12} for all the 10 different parameter configuration. The true value of the parameter (equal for the case A and B) is 0.95 for $\psi_2 = 0.1$, 0.84 for $\psi_2 = 0.3$, 0.70 for $\psi_2 = 0.5$, 0.55 for $\psi_2 = 0.7$, 0.32 for $\psi_2 = 0.9$. We note that estimates become stable with a few dozen repetitions, so observing these graphs we could fix $N = 100$. Then we focused on the standard error of the estimates. In Figure 3.4, we report the simulated mean standard error estimates of regression coefficient β_{12} for all the 10 different parameter configuration. We can note that we need more iterations, and so more samples, to have stable standard error means. Considering, therefore, the stability of the estimates and the trend of the standard errors, we decided to set the number of samples (N) equal to 500, for each parameter set. The size of samples, N greatly influences the computing time, for this reason we tried to keep it relatively low, while the sample size n, had a lower impact on it.

3.3.1 The Data Generation Process

The data have been generate with M*plus*, following the model described in equations (3.1) and (3.2). The parameters have been set as presented in the previous paragraph. For the generation procedure, we had to decide whether to use a probit or logit model (Agresti, 1996, Muthén and Muthén, 2007). To understand which to choose, we made another small simulation. We generated some samples with both models, and then we verified the results, obtained using different software: M*plus*, "*eRm*" package of *R*, and *Winstep* (Linacre, 2003). We found that with the probit model, we obtained more consistent results with the different softwares. For a binary dependent variable, the probit regression model expresses the probability of response

as:

$$\begin{aligned} P(y_{ij}=1|\theta_i) &= F(a_0+a_1\theta_i) \\ &= F(-\tau_j+a_1\theta_i), \\ P(y_{ij}=0|\theta_i) &= 1-P(y_{ij}=1|\theta_i), \end{aligned}$$

where F is the standard normal distribution function, a_0 is the probit regression intercept, a_1 is the probit regression slope, τ_j is the probit threshold (in this case $\tau_j = -a_0$). For the ordered categorical dependant variables with 5 categories, the probability of response for the probit regression model is:

$$\begin{aligned} P(y_{ij}=0|\theta_i) &= F(\tau_{j1}-a_1\theta_i) \\ P(y_{ij}=1|\theta_i) &= F(\tau_{j2}-a_1\theta_i)-F(\tau_{j1}-a_1\theta_i) \\ P(y_{ij}=2|\theta_i) &= F(\tau_{j3}-a_1\theta_i)-F(\tau_{j2}-a_1\theta_i) \\ P(y_{ij}=3|\theta_i) &= F(\tau_{j4}-a_1\theta_i)-F(\tau_{j3}-a_1\theta_i) \\ P(y_{ij}=4|\theta_i) &= F(-\tau_{j4}+a_1\theta_i), \end{aligned}$$

where τ_{jk} are the probit thresholds.

We started generating multivariate normal data for the independent variables in the model (θ_1, θ_3, Δ_j). Then the data for the continuous dependent variable, θ_2, 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, τ_{hk}, and the item difficulties, δ_j, have been chosen to obtain items with different response distributions (positively asymmetric, symmetric and asymmetric negative). We tried to maintain these distributions in all sets of parameters generated (Figure 3.5).

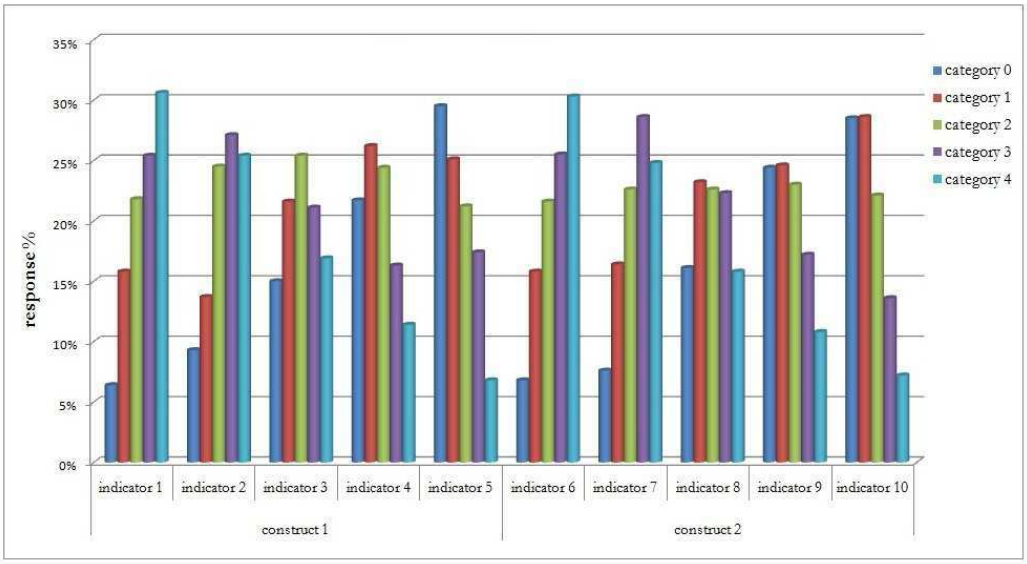


Figure 3.5. The frequency distribution for the item response categories of the 10 items of the first scenario, case A, $\psi_2 = 0.1$.

3.4 The Simulation Results

In this Section, we will present the key findings from the simulation study. All reported results refer to the estimated regression coefficients β_{12} and β_{32}, which estimate the dependence relationships between the latent variables, the final goal of many of the socio-economic studies that we mentioned earlier. To compare the results obtained with the 4 estimation methods, we evaluate:

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

We choose to use relative indices, that is they have been divided by the actual value of the parameter, to allow a correct comparison in the several presented cases (important forethought because the regression coefficient takes values in $[0.1, 0.9]$).

3.4.1 First Scenario

We start the result analysis from the first scenario, where we have two latent variables, the second one (θ_2) dependent from the first construct (θ_1).

In Figure 3.6, we see the relative bias, $RB(\hat{\beta}_{12})$, for the regression coefficient β_{12}. In boxes (**a**) and (**b**), we find the results for the cases in which the independent construct has 5 or 10 items, respectively. We can immediately observe that the results of **RSM-LRM** method show a strong negative bias, consistent with the theory of measurement errors (Section 1.3.1 at page 7). In boxes (**c**) and (**d**), we brought the same results of the two previous ones, by removing those of the fourth method. **SEMstd** and **SEMirt** methods show a distortion of reduced entity (in absolute value less than 1%). It is interesting to note that the broken line of the two One-step procedures are practically overlapping, though they represent, conceptually, two very different methods. Indeed, we recall that the method **SEMirt** corresponds to the data generator model, while the less constrained **SEMstd** method is a traditional SEM, which does not take into account the IRT structure. The bias of the Two-step procedure **RSM-LRMme** increases as ψ_2, the variance of structural error term ζ_2, increases.

It is also interesting to note that, in case B, the distortion of the **RSM-LRMme** method is less than the case with few indicators for the independent latent variable, consistent with IRT.

In Figure 3.7, we see the relative standard errors, $RSE(\hat{\beta}_{12})$, for the regression coefficient β_{12}. The **RSM-LRM** method shows relative standard errors lower than other methods. The broken line of the two One-step procedures, again, are practically overlapping and they show the same trend; while the Two-step procedure **RSM-LRMme** is the less precise. For all the 4 methods, the standard errors increase as ψ_2, the variance of structural error term ζ_2, increases.

In Figure 3.8, we see the relative $RMSE(\hat{\beta}_{12})$ for the estimates of the regression coefficient β_{12}. From this graph, we note the importance of considering, in the estimation procedure, the measurement errors that affect the variables.

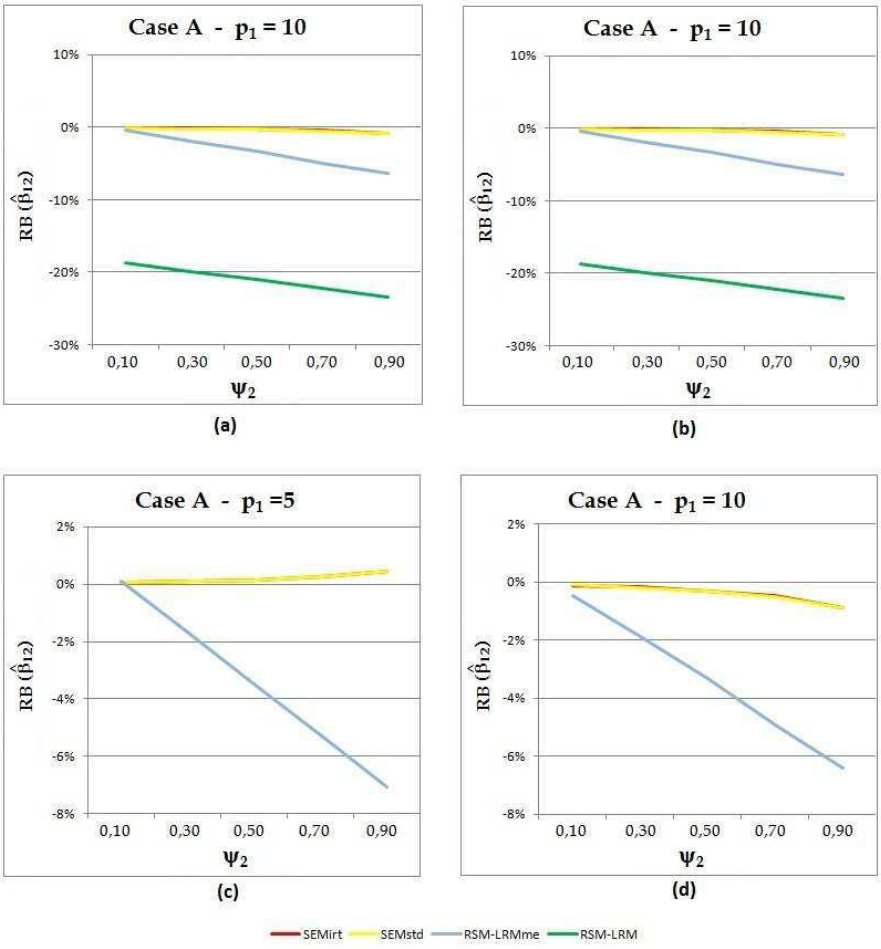


Figure 3.6. Simulation Results: the Relative Bias (RB) for the estimates of the regression coefficient β_{12} in the first scenario

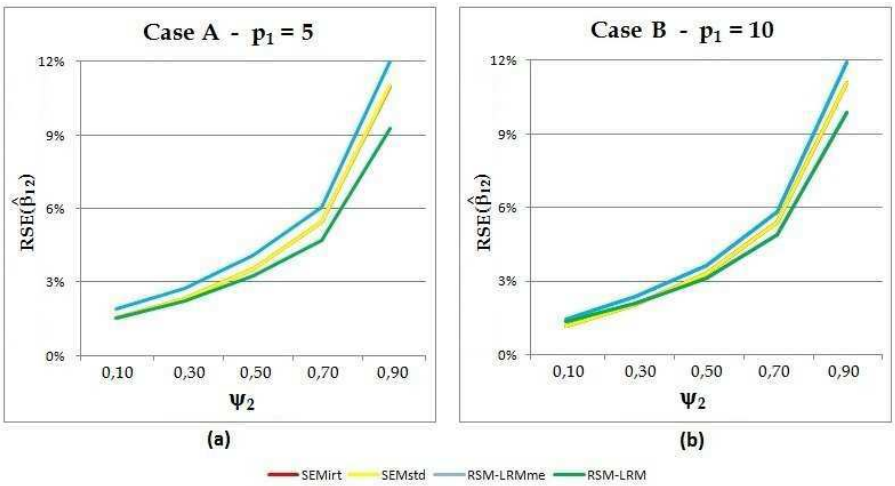


Figure 3.7. Simulation Results: the Relative Standard Error (RSE) for the estimates of the regression coefficient β_{12} in the first scenario

In fact, the only method that ignored the problem, the **RSM-LRM**, presents a very high relative $RMSE(\hat{\beta}_{12})$ (up to 10 times the other method values). It is due to the strong bias, not sufficiently compensated by the good accuracy in estimating. The broken line of the two One-step procedures, **SEMirt** and **SEMstd**, are practically overlapping and they have the lowest RMSE, less than 10%, except in the case of maximum relative variance, where is nearly 12%. The **SEM-LRMme** method has a slightly higher $RMSE(\hat{\beta}_{12})$, although the discrepancy with the One-step procedure is never more than 3 percentage points. For example, in the case B with $\psi_2 = 0.9$, for **SEMirt** we have $RMSE(\hat{\beta}_{12}) = 11.1\%$, while for the **RSM-LRMme** equal it is equal to 13.6%.

Similar considerations can be drawn observing the Boxplot chart in Figure 3.9. Boxplots (also known as a box-and-whisker diagrams) display the differences between populations without making any assumptions of the underlying statistical distribution. The ends of the whiskers represent the minimum and maximum of the sample data; the dashed red line indicates the true value of the estimated parameter. This graph shows very well the relationship between accuracy of the estimate and the number of indicators used for the measures. We actually note that all estimates of case B present

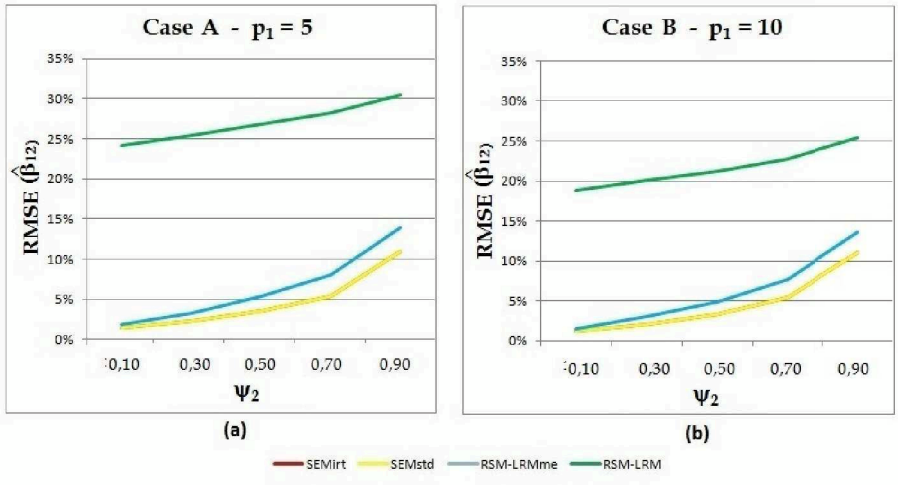


Figure 3.8. Simulation Results: the Relative Root Mean Square Error (RMSE) for the estimates of the regression coefficient β_{12} in the first scenario

smaller box plots, and they are less dispersed around their median. Furthermore, this representation highlights the strong distortion of **SEM-LRM** method. To simplicity, we present only one of five available boxplot charts (one for each level of ψ_2), the others have a similar trend.

Focusing for a moment on the Two-step procedure, we can analyze the reliability indices of the estimated measures. We compared the values of Cronbach's alpha and the Rasch Person Reliability Index (RPRI), proposed in the chapter 2 (Section 2.4.4 at page 50), and we have seen that they are perfectly consistent, as evidenced the Figure 3.10 **(a)**. We can just multiply the RPRI index by a factor g (in this case $g\epsilon[1.2, 1.3]$) to obtain the corresponding value of α. Furthermore, it is important to observe that the index values for the two measures, Y_1^* and Y_2^*, estimated at the first stage of the Two-step approach, are always above the threshold 0.7, which is referred to as the minimum acceptable level for Cronbach's alpha (Nunnally and Bernstein, 1994). It is also very interesting to note that Y_1^*, which is associated with the construct measured by 10 indicators, presents reliability indices significantly higher than Y_2^*. It is also worth emphasizing that this increase is observed, similarly, in both of the indices.

In Figure 3.10 **(b)** and **(c)**, we present Cronbach's alpha index values,

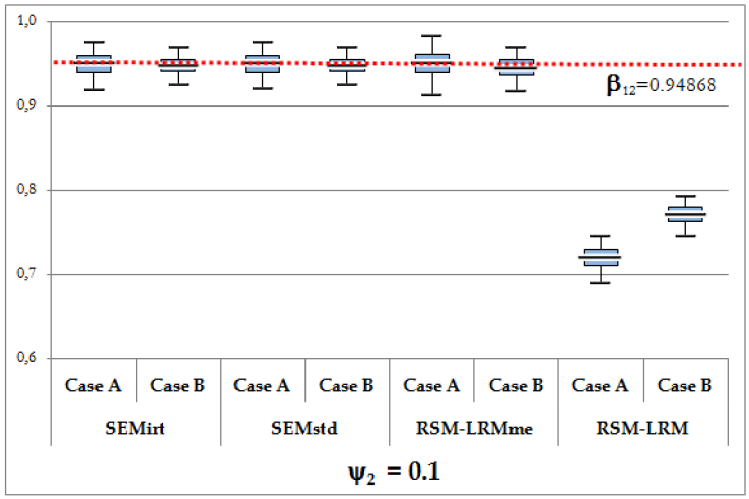
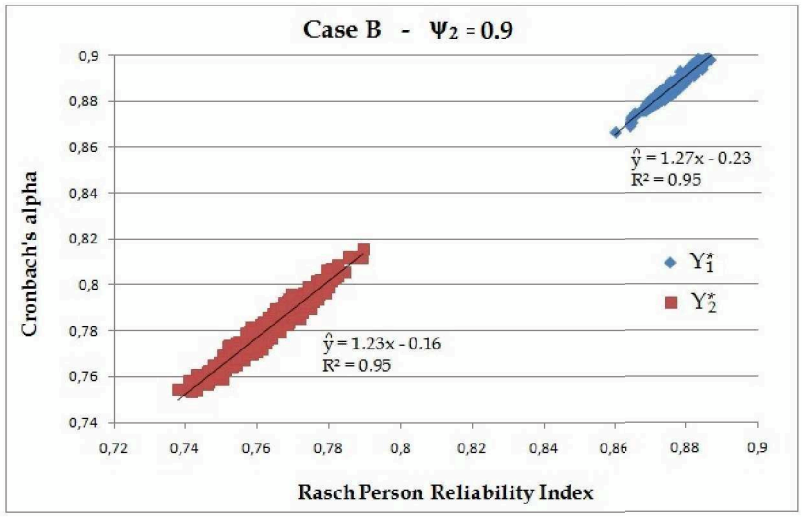


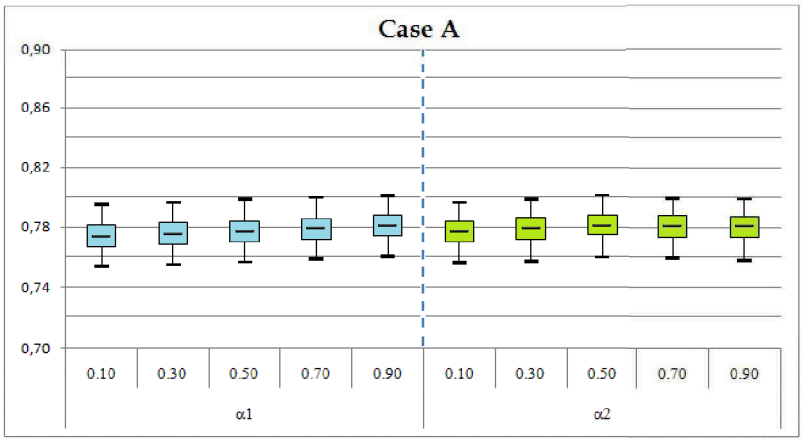
Figure 3.9. Simulation Results: the boxplot chart for the estimate of the regression coefficient b_{12} in the first scenario

more used in literature, for the two measures, Y_1^* and Y_2^*. This graph, again, clearly shows the increase in reliability due to the increase of indicators for the first construct. In fact, the 5 Boxplots related to the independent latent variable of case B clearly show that the reliability indices in these samples are significantly higher than others; their α_1 values increased by almost 10 percentage points compared to that of the measures with only 5 indicators. From this graph, we can deduce that the weakening of dependence relationship between the two latent variables, does not affect the estimation of measurement associated with each latent variable.

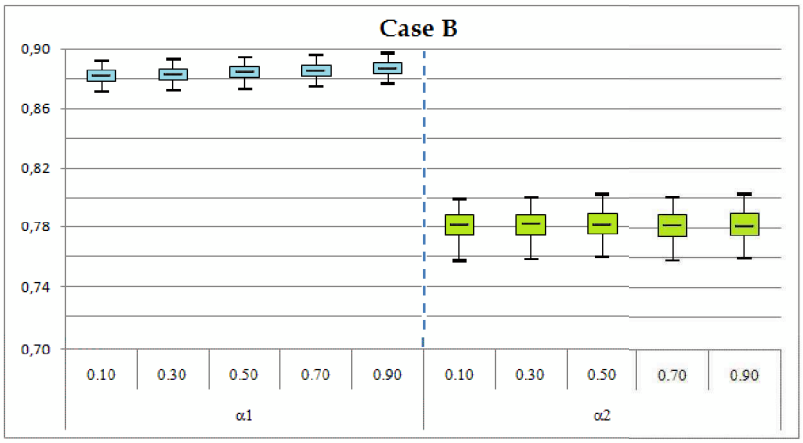
This reliability analysis, in addition to a series of other considerations about, for example, the unidimensionality, the correct order of categories and the differential item functioning (Hambleton et al. 1991; Roju et al. 1995; Emons et al. 2005; Zumbo, 2007), are possible only with the Two-step procedure, that separates the estimation of the measures from the estimation of the construct relationships. It is a very important step in the latent variable analysis, which has not an analogue into the One-step procedure, where we can only verify the goodness of fit, but not the reliability of the constructs (in IRMs terms).



(a)



(b)



(c)

Figure 3.10. Simulation Results first scenario: **(a)** comparison between the RPRI values and the Cronbach's alpha values for the two measures Y_1^* and Y_2^*, **(b)** and **(c)** the boxplot chart for the Cronbach's alpha indices for the two measures Y_1^* and Y_2^*

3.4.2 Second Scenario

The second scenario presents three latent variables, where θ_2 depends from θ_1 and θ_3. The results are similar to those already presented for the first scenario, we will briefly summarize the most salient.

In Figure 3.11, we see the relative bias for the regression coefficients β_{12} and β_{32}, case A $(\frac{3}{4}; \frac{1}{4})$, where $\beta_{12} = 3\beta_{32}$. As for the first scenario, we can immediately observe that the results of **RSM-LRM** show a strong negative bias, consistent with the theory of measurement errors (1.7 at page 7). In boxes **(c)** e **(d)**, we brought the same results of boxes **(a)** e **(b)**, by removing those of the fourth method. 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 Two-step procedure **RSM-LRMme** increases as ψ_2 increases, but in absolute value is always lower than 10%.

In Figure 3.12, we see the relative standard error for the estimates of the regression coefficients β_{12} and β_{32}. The **RSM-LRM** method shows $RSE(\hat{\beta}_{12})$ and $RSE(\hat{\beta}_{32})$ lower than other procedures, even if they are all very close. While for the case $(\frac{1}{2}; \frac{1}{2})$ the estimates of two different coefficients have standard errors of equal magnitude, we note that for the case $(\frac{3}{4}; \frac{1}{4})$, with $\psi_2 = 0.9$, the relative standard errors of $\hat{\beta}_{12}$ are very high, between 30% and 40% of the parameter value. For all the 4 methods, the standard errors increase as ψ_2 increases.

The relative $RMSE(\hat{\beta}_{12})$, for the estimates of the regression coefficients β_{12} and β_{32} are shown in table 3.1. In the last rows of the table, we reported the actual values of the regression coefficients. Consistent with what is already shown in the first scenario, we observe that **RSM-LRM**, ignoring the presence of measurement errors, presents a very high RMSE (up to 3 times that of other methods). It is due to strong bias already seen in previous graphs, not sufficiently compensated by the good accuracy in estimating. The two One-step procedures, **SEMirt** and **SEMstd**, have the lowest relative RMSE, even if the **SEM-LRMme** relative RMSE values are very close to the previous, only from 2 to 4 percentage points more. The very high values (above 30%), occurred when $\psi_2 = 0.90$, $\hat{\beta}_{12}$ $(\frac{1}{4}; \frac{3}{4})$ and

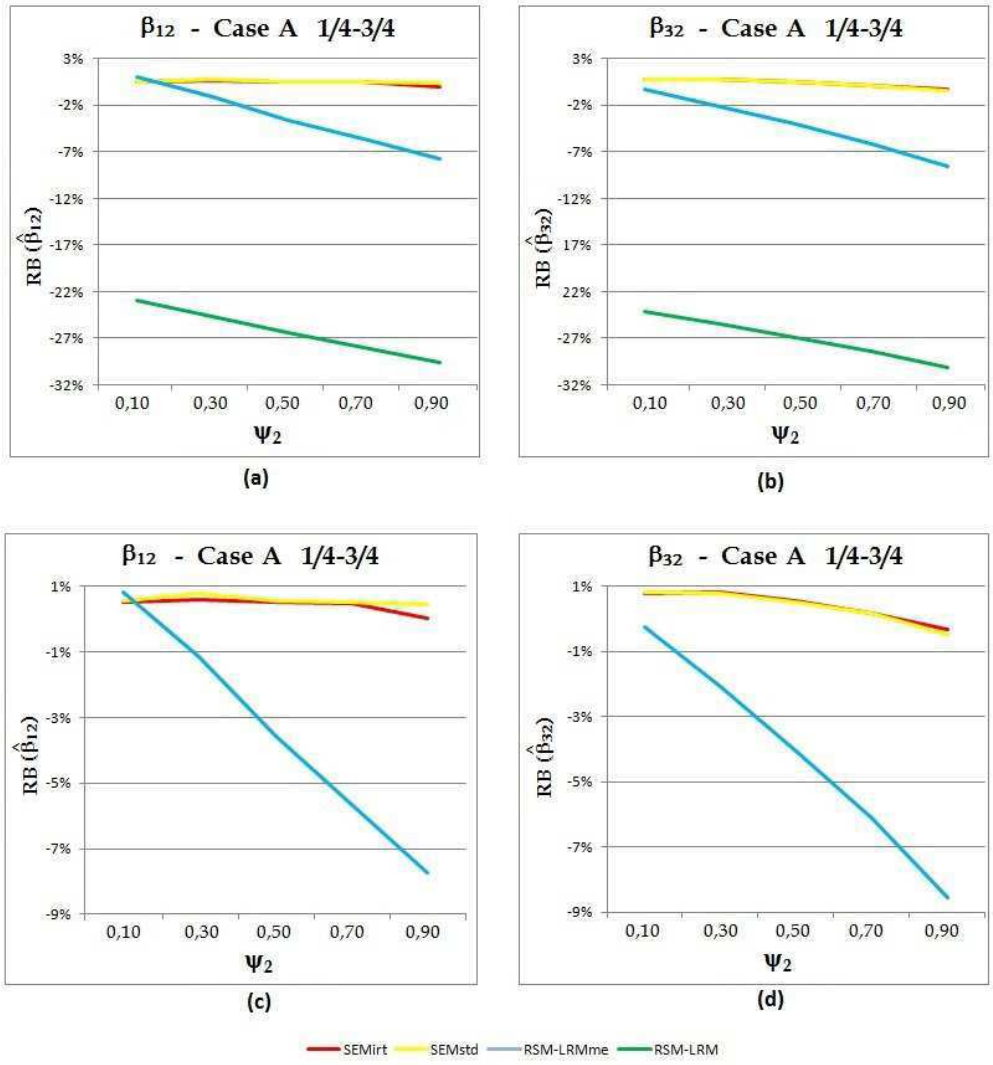


Figure 3.11. Simulation results: the Relative Bias (RB) for the estimates of the regression coefficients β_{12} and β_{32} in the second scenario case A $(\frac{3}{4}; \frac{1}{4})$, where β_{12} is the 75% of κ and β_{32} is the 25% of κ

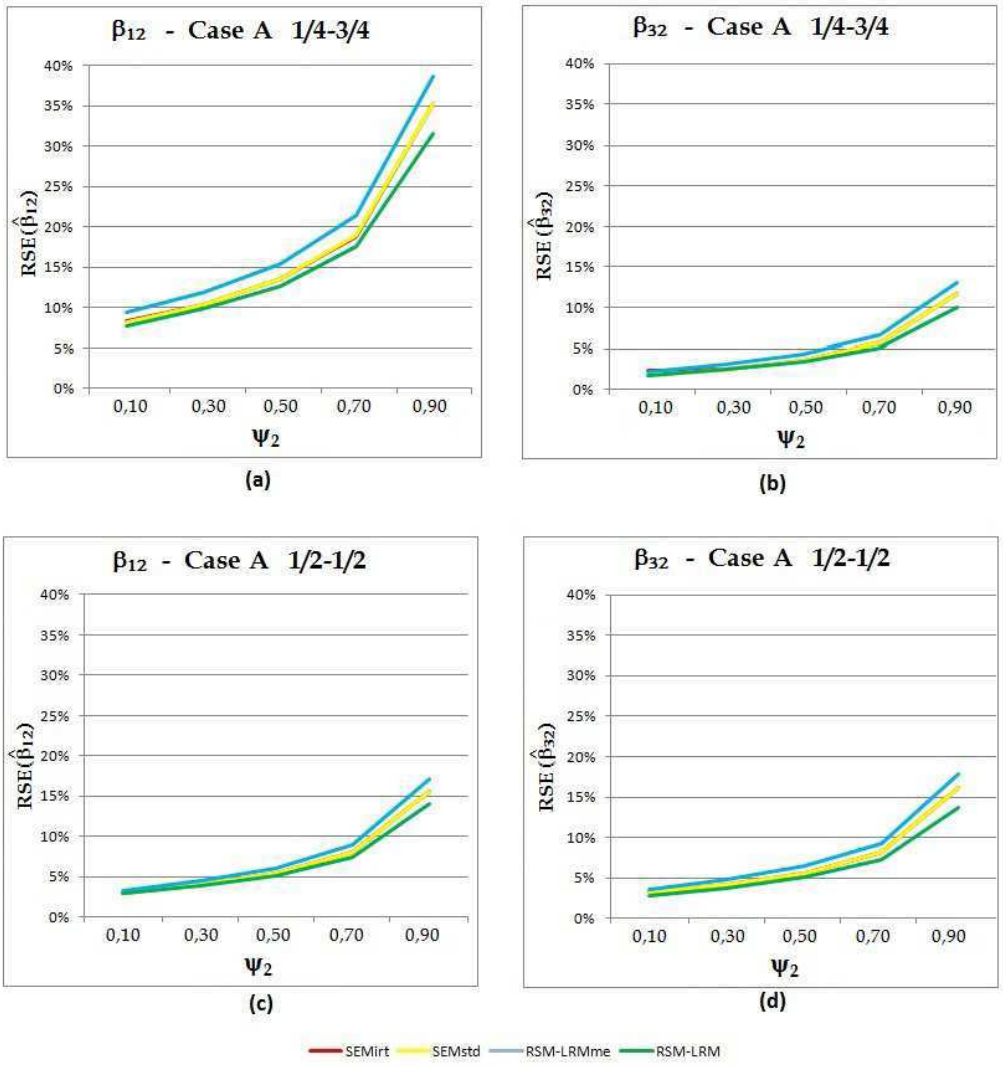


Figure 3.12. Simulation results: the Relative Standard Error (RSE) for estimates of the regression coefficients β_{12} and β_{32} in the second scenario case A

Table 3.1. Simulation results: the relative Root Mean Square Error (in %) for the estimate of the regression coefficients β_{12} and β_{32}, in the second scenario case B

			0.10	0.30	0.50	0.70	0.90
$\hat{\beta}_{12}$	1/4-3/4	SEMirt	9.32	10.40	13.56	18.88	35.26
		SEMstd	8.24	10.42	13.57	18.93	35.30
		RSM-LRMme	9.65	11.96	15.43	21.76	39.03
		RSM-LRM	17.65	20.07	23.13	27.52	39.17
	1/2-1/2	SEMirt	3.02	4.12	5.48	8.07	15.56
		SEMstd	3.02	4.11	5.48	8.09	15.67
		RSM-LRMme	3.39	4.57	6.49	9.75	17.97
		RSM-LRM	17.03	18.71	20.44	22.59	26.55
	3/4-1/4	SEMirt	1.39	2.43	3.54	5.53	11.32
		SEMstd	1.41	2.42	3.53	5.53	11.34
		RSM-LRMme	1.69	3.10	4.93	7.64	13.78
		RSM-LRM	18.31	19.50	20.76	22.24	25.05
$\hat{\beta}_{32}$	1/4-3/4	SEMirt	4.79	2.78	3.69	5.76	11.74
		SEMstd	1.84	2.78	3.66	5.74	11.70
		RSM-LRMme	2.18	3.70	5.84	8.88	15.07
		RSM-LRM	24.19	25.53	27.01	28.71	31.07
	1/2-1/2	SEMirt	3.28	4.32	5.66	8.23	16.13
		SEMstd	3.28	4.30	5.63	8.21	16.18
		RSM-LRMme	3.73	5.39	7.70	10.98	19.37
		RSM-LRM	24.02	25.54	27.18	28.95	32.40
	3/4-1/4	SEMirt	7.62	10.36	13.69	19.45	37.43
		SEMstd	7.62	10.33	13.65	19.39	37.36
		RSM-LRMme	8.72	11.68	15.79	21.96	40.88
		RSM-LRM	24.46	26.69	29.08	32.56	42.31
True value parameter	β_{12}	1/4-3/4	0.30	0.26	0.22	0.17	0.10
		1/2-/12	0.67	0.59	0.50	0.39	0.22
		3/4-1/4	0.90	0.79	0.67	0.52	0.30
	β_{32}	1/4-3/4	0.90	0.79	0.67	0.52	0.30
		1/2-/12	0.67	0.59	0.50	0.39	0.22
		3/4-1/4	0.30	0.26	0.22	0.17	0.10

$\hat{\beta}_{32}$ $(\frac{3}{4}; \frac{1}{4})$, are due to the fact that the index is divided by 0.1 (actual value of the parameters β_{12} and β_{32}, respectively).

In Figure 3.13, we have the Boxplot charts, that underly, again, the strong distortion of **SEM-LRM** method.

Finally, we pay attention on the reliability indices for the measures of Two-step procedure As in the first scenario, the two indices, RPRI and Cronbach's alpha, are perfectly consistent. We can just multiply the RPRI index by a factor g for computing the value of α. We present the Cronbach's alpha results, which is certainly more used in literature. In Figure 3.14, we can see the index for the three measures, Y_1^* Y_2^* and Y_3^*, estimated at the first stage of the Two-step approach. We note that all the index values are significant because they are greater than 0.7. Furthermore the measures built across 10 indicators are characterized by an index of reliability far superior to the others.

Time computer processing

To conclude the analysis of results, we also talk about the computer processing time required by the different methods. For the first scenario, the method **SEMstd** takes about 7 hours for each parameter configuration of case A and more than 12 hours for each set of parameters of case B, a total of about 100 hours of elaboration. **SEMstd** method takes about 15 hours for each parameter configuration of case A and more than 27 hours for each parameter configuration of case B, a total of about 230 hours of elaboration. **RSM-LRMme** method took about 16 hours for all parameter configurations of case A and 1 day for all configurations of case B, in total about 40 hours of elaboration. **RSM-LRM** method, actually based on information estimated from the first step the other method, took 5 hours total, for both Case A and B, to perform only the second step. Thus, for the first scenario, we note that the procedures Two-step are considerably faster than One-step procedures.

This difference becomes much more significant with the second scenario. **SEMstd** method requires approximately 8 hours for each parameter

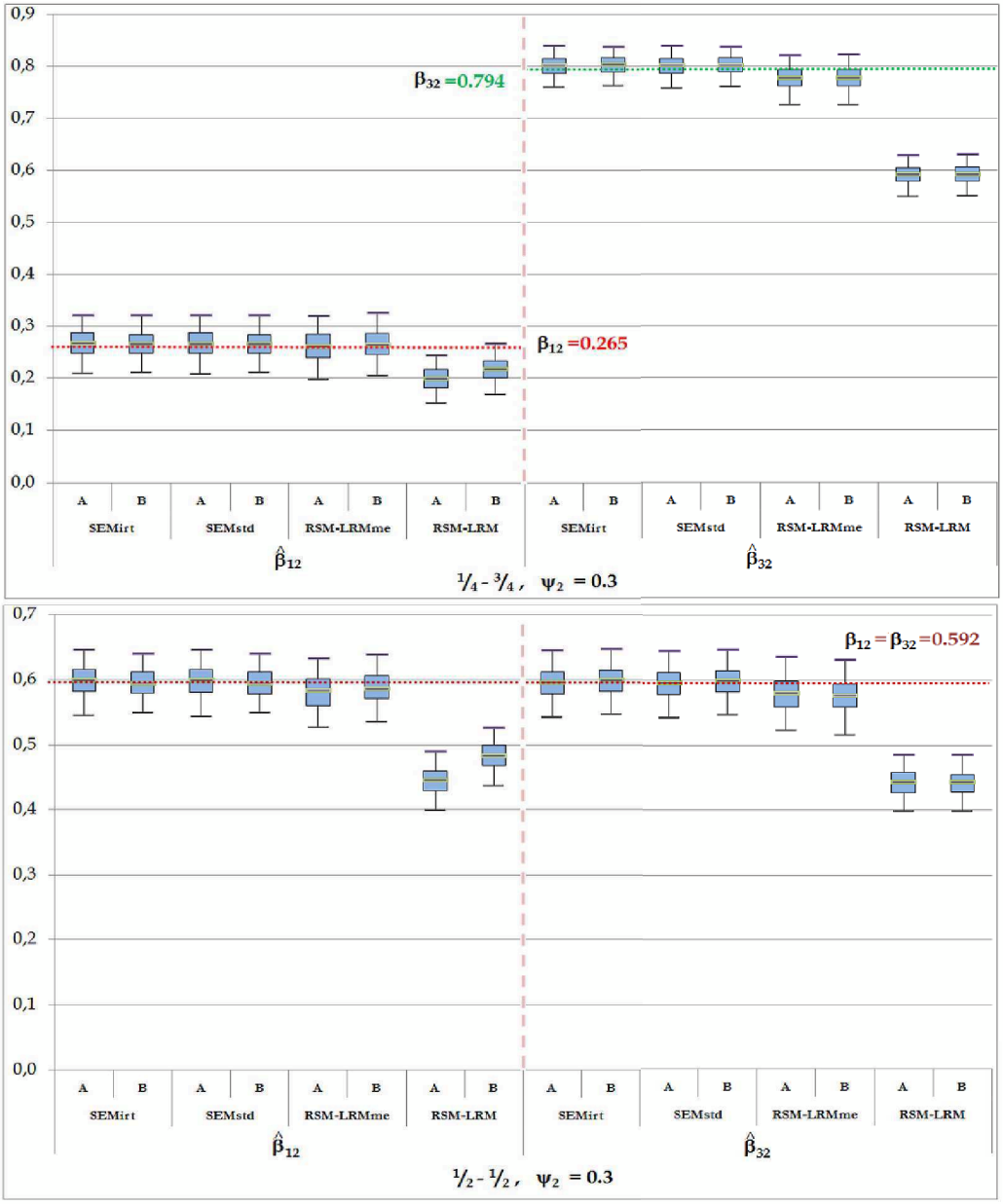


Figure 3.13. Simulation results: the boxplot charts for the two regression coefficients β_{12} and β_{32} in the second scenario, case $\psi_2 = 0.3$, with distribution $(\frac{1}{4}; \frac{3}{4})$ and $(\frac{1}{2}; \frac{1}{2})$

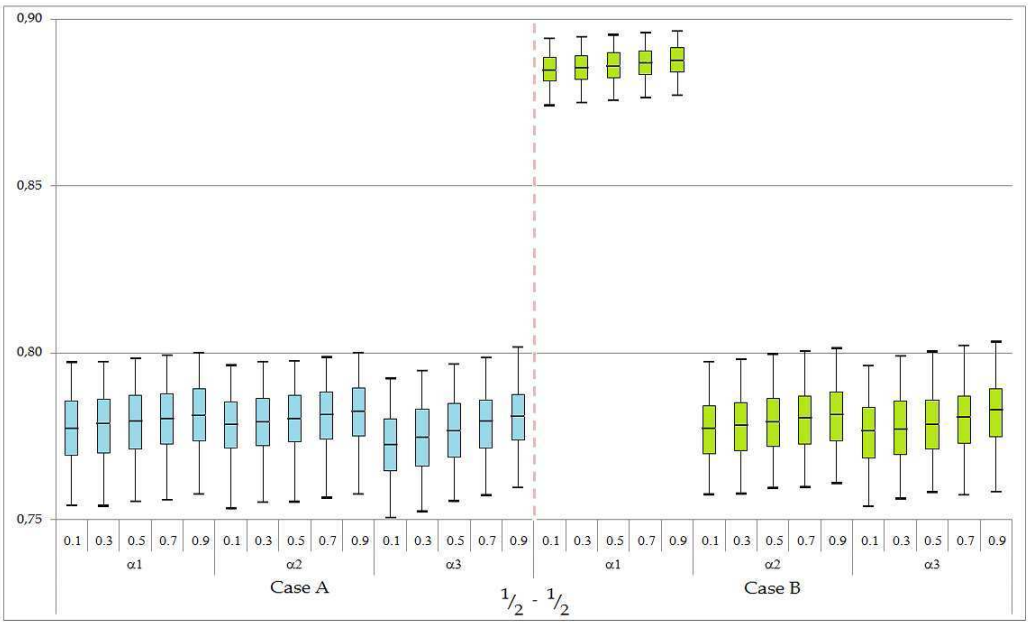


Figure 3.14. Simulation results: the boxplot chart for the Cronbach's alpha indices for the three measure Y_1^*, Y_2^* and Y_3^* in the second scenario, with distribution $(\frac{1}{2}, \frac{1}{2})$

configuration of case A and more than 19 hours for each set of parameters of case B, a total of about 380 hours of elaboration. **SEMstd** method takes about 21 hours for each parameter configuration of case A and more than 35 hours for each set of parameters of case B, a total of about 750 hours of elaboration. **RSM-LRMme** method took about 30 hours for all parameter configurations of case A and two days for all configurations of case B, in total about 40 hours of processing. **RSM-SEM** took, to perform only the second step, a total of 2 hours, including both the cases A and B.

These processing times are the averages of the elaborations carried out on faster computers, but during the simulation process we had to use less powerful machines which have considerably slowed down the analysis. Overall, the simulations with the 35 presented configurations took more than three months, at which we should add the time for the simulations carried out to determine the parameter values presented in Section 3.3.

Conclusion and Future Research

A first consideration is about the **RSM-LRM** method. Although at times the standard linear regression is used also with variables affected by measurement errors, our simulation showed that the bias of estimator for the parameter of interest is very strong. Precisely for this reason, in our conclusions, we consider, for the Two-step procedure, only the **RSM-LRMme** method.

Remembering that one of our goals was to implement a Two-step procedure efficient and precise, we focus the attention on the Root Mean Square Error index (that combines an assessment of bias and efficiency). From the reported results in table 3.2, study, referred only to the **SEMstd** and the **RSM-LRMme** methods, we can deduce that, consistent with the theory proposed in Chapter 2, the Two-step procedure has a slight distortion and a loss of efficiency, but its estimates are coherent with that providing by the One-step procedure and often the difference with them is really very small. The difference between the two procedures is just few percentage points (up to 4.3 in the second scenario, for the case A, β_{12}, $(\frac{1}{4}, \frac{3}{4})$). It is a very interesting result, that provides a useful tool for future analysis starting to real data.

We have repeatedly stressed the advantages of IRM in terms of greater flexibility of analysis and possibility of verification of 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.

Table 3.2. Simulation results: A comparison of all RMSE indices for all the different parameter configurations for the **SEMstd** and the **RSM-LRMme** method

					0.10	0.30	0.50	0.70	0.90
First scenario	A			SEMstd	1.50	2.30	3.52	5.45	11.02
				RSM-LRMme	1.88	3.23	5.39	8.03	13.93
	B			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	A	β_{12}	1/4-3/4	SEMstd	8.53	10.87	14.24	20.12	37.09
				RSM-LRMme	9.69	12.16	16.26	23.07	41.32
			1/2-1/2	SEMstd	3.44	4.58	5.98	8.59	16.56
				RSM-LRMme	3.85	5.24	7.65	11.27	19.95
		β_{12}	1/4-3/4	SEMstd	1.83	2.74	3.86	5.91	11.92
				RSM-LRMme	2.14	3.71	6.01	9.08	15.91
			1/2-1/2	SEMstd	3.34	4.46	5.83	8.44	16.37
				RSM-LRMme	3.71	5.42	7.80	11.49	20.30
	B	β_{12}	1/4-3/4	SEMstd	8.24	10.42	13.57	18.93	35.30
				RSM-LRMme	9.65	11.96	15.43	21.76	39.03
			1/2-1/2	SEMstd	3.02	4.11	5.48	8.09	15.67
				RSM-LRMme	3.39	4.57	6.49	9.75	17.97
			3/4-1/4	SEMstd	1.41	2.42	3.53	5.53	11.34
				RSM-LRMme	1.69	3.10	4.93	7.64	13.78
		β_{12}	1/4-3/4	SEMstd	1.84	2.78	3.66	5.74	11.70
				RSM-LRMme	2.18	3.70	5.84	8.88	15.07
			1/2-1/2	SEMstd	3.28	4.30	5.63	8.21	16.18
				RSM-LRMme	3.73	5.39	7.70	10.98	19.37
			3/4-1/4	SEMstd	7.62	10.33	13.65	19.39	37.36
				RSM-LRMme	8.72	11.68	15.79	21.96	40.88

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

Even if the second scenario represents already a degree of complexity present in many real researches, the obtained results, providing a good indication for future research in this area, are not generalizable to more complex analysis (for example with three or more regressors, or with correlated regressors).

Another finding obtained by simulated data is the increase of reliability in the case of the latent variables measured by a greater number of indicators. As already mentioned, it is a result known in the literature for the IRMs, but our simulations also showed a similar result for the SEM. From Table 3.2, it can be noted that, for the first construct, the value of RMSE for the case B is in all cases lower than the analogous of case A also for the One-step procedure. Therefore, we can say that the introduction of additional indicator variables, although affected by measurement errors, improves the model goodness.

With reference to the analysis of reliability indices, moreover, the simulation allowed us to compare the performance of two indices used in this context: Cronbach's alpha and Rasch Person Reliability Index (RPRI). The data showed that these indices followed a similar pattern, thus providing similar indications. In our **RSM-LRMme** method, we decide to use the RPRI as it is the natural index of reliability under the IRMs (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, which we give an estimate of the variance of their errors. For logical consistency, these estimates are represented by the RPRI values obtained in the previous step.

All the above considerations do not take into account computer elaboration times. We did not consider them, because it is a typical

problem of simulation studies, not present in real cases, where we do not have to test hundreds of samples, but at most a few dozen. The procedure for Two-step is the fastest, but about a dozen samples the difference is in the order, at most, of a couple of hours.

Our simulation study, although provided many different cases, should be considered as a baseline for future analysis. We have already begun to study the case with measurement errors of different sizes, but in this case we must be careful not to generate data that present too low reliability indices.

Another interesting research cue will be the extension to multilevel cases. For SEM, there is already an extensive literature on the subject, while for the IRT, although some models have been proposed, there is not still a substantial theoretical basis.

Finally, because as we said earlier, this research draws upon a real problem, we would apply the estimation methods to a real case, obviously having to deal with any problems that this entails (eg missing data, unreliable measures, incorrect orders in response categories).

Appendices

Appendix A

Basic Statistical Concepts

TO entering the analysis of SEMs, it could be useful to recall briefly some statistical prerequisites: correlation, regression, data preparation and screening.

A.1 Pearson Correlation

The Pearson correlation

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \qquad \rho_{xy} \, \epsilon \, [-1, 1],$$

estimates the degree of linear association between two continuous variables X and Y, where σ_x, σ_y are their standard deviations and σ_{xy} is their covariance. If $\rho_{xy} > 0$, the variables X and Y are said to be directly (or positively) correlated, when $\rho_{xy} = 0$ there is no linear relation (but there could be a curvilinear association), if $\rho_{xy} < 0$, X and Y are inversely (or negatively) correlated. ρ_{xy} could be about zero not only when the relation between X and Y is nonlinear, but also when the variance of either X or Y is relatively narrow, when the shapes of the frequency distributions of the two variables are very different, or when the reliability of the scores on either X or Y is about zero.

The variable means don't affect ρ_{xy} or σ_{xy}, so if we are interested to changes in means, we have to choose a SEM method that can include them

in the analysis.

A.2 Regression

We start from the case of only two continuous variables, the predictor X and the dependent variable Y. Based on their association, the regression equation is

$$\hat{Y} = b_0 + b_x X.$$

b_x is the unstandardized regression coefficient and it indicates the predicted difference on Y given a 1-point increase on X. The residuals $(Y - \hat{Y})$ reflect the part of Y that cannot be explained by X, so they are uncorrelated with X. The coefficient b_x is related to the standard deviations of X and Y and the pearson correlation index ρ_{xy}:

$$b_x = \rho_{xy} \frac{\sigma_y}{\sigma_x}.$$

From the equation above, we can understand that if we have two standardized variables, their correlation is also the standardized regression coefficient, It indicates the expected difference on Y in standard deviation units, given an increase on X of one full standard deviation.

We can extend this reasoning to the case of multiple regression analysis, with $p, (p \geq 2)$ predictors:

$$\hat{Y} = \alpha + b_{x_1} X_1 + \ldots + b_{x_p} X_p,$$

where $(b_{x_1}, \ldots, b_{x_p})$ indicates the expected raw score difference in Y given a difference of a single point in one predictor while we are controlling for the others. The multiple regression equation for standardized variables is

$$\hat{Z}_y = \beta_{x_1} Z_{x_1} + \ldots + \beta_{x_p} Z_{x_p},$$

where $(Z_y, Z_{x_1}, \ldots, Z_{x_p})$ are the standardized variables of $(Y, X_1, \ldots, X_p)$ and $(\beta_{x_1}, \ldots, \beta_{x_p})$ are the standardized regression coefficients, also called *beta*

weights.

Given two predictors, the formulas for the beta weights are:

$$\beta_{x_1} = \frac{\rho_{yx_1} - \rho_{yx_2}\rho_{x_1x_2}}{1 - \rho^2_{x_1x_2}} \qquad \beta_{x_2} = \frac{\rho_{yx_2} - \rho_{yx_1}\rho_{x_1x_2}}{1 - \rho^2_{x_1x_2}}, \tag{A.1}$$

where ρ_{yx_1}, ρ_{yx_2} and $\rho_{x_1x_2}$ are the bivariate correlations among the dependent variable and the predictors. Because beta weights are adjusted for intercorrelations among the predictors, their absolute values are usually less than those of the corresponding bivariate correlations of the predictors with the dependent variable. Note that if the predictors are independent ($\rho^2_{x_1x_2} = 0$), then each beta weight equals the corresponding bivariate correlation, because there is no adjustment for correlated predictors. The relation between unstandardized and standardized regression coefficients in multiple regression is

$$b_{x_1} = \beta_{x_1}\frac{\sigma_y}{\sigma_{x_1}} \qquad b_{x_2} = \beta_{x_2}\frac{\sigma_y}{\sigma_{x_2}} \ .$$

A.3 Data Preparation and Screening

Basically all SEM methods can analyze either a raw data file or a matrix summary of the data. The following issues should be considered in choosing between a raw data file and a matrix summary as program input:

- We have to use raw data file when a nonnormal data are analyzed with an estimation method that assumes normality but test statistics are calculated that correct for nonnormality or when we use a special estimation method that does not assume normal distributions or accommodates cases with missing observations.

- Matrix input offers a potential economy over raw data files especially for large samples.

- Sometimes we want to create a correlation (or covariance) matrix using theory or results from a meta-analysis, so we don't have raw data, but only a data matrix.

Before a raw data file or a matrix summary is created, the original data should be carefully screened for the following problems: normality of the data, missing observations, multicollinearity, model identification.

The most widely used estimation methods in SEM assume multivariate normality, which means that all the univariate distributions are normal, the joint distribution of any pair of the variables is bivariate normal and all bivariate scatterplots are linear and homoscedastic. Fortunately, many instances of multivariate nonnormality are detectable through inspection of univariate distribution.When we talk about normality of the data we must first evaluate skew and kurtosis (DeCarlo, 1997) of the data univariate distribution. If the shape of a unimodal distribution is asymmetrical about its mean, we have positive skew if the most scores are below the mean, or negative skew otherwise. For a unimodal, symmetrical distribution, we talk about positive kurtosis if we have heavier tails and a higher peak and about negative kurtosis in the opposite case. One way to deal with univariate nonnormality is with transformations, for example square root, logarithmic, and inverse functions. Another problem to be addressed during the preliminary analysis of data is the case of scores that appear to deviate markedly from other members of the sample, called outliers.

It could happen that the dataset has missing observations, if they are not systematic then we can talk about ignorable data loss patterns missing at random (MAR), when missing observations on some variable X differ from the observed scores on that variable only by chance, or missing completely at random (MCAR), when the presence versus absence of data on X is unrelated to any other variable (Allison, 2003). Many authors have addressed this problem and several solutions have been proposed in the literature, in different contexts and for different types of data (Jones, 1996, Little and Rubin, 2002, Zarate et al. 2007).

Another cause of singular covariance matrices is multicollinearity, which occurs when intercorrelations among some variables are so high that certain mathematical operations are either impossible or unstable because some denominators are close to zero. Multicollinearity can occur because more variables actually measure the same thing. It is easy to spot pairwise

multicollinearity simply by inspecting the correlation matrix, it is more difficult to detect multicollinearity among three or more variables. One method is to calculate a squared multiple correlation, R^2_{smc}, between each variable and all the rest. A related statistic is tolerance ($TOL = 1 - R^2_{smc}$) and it indicates the proportion of total standardized variance that is unique, that is not explained by all the other variables. Another statistic is the Variance Inflation Factor (VIF), equals to the ratio of the total standardized variance to unique variance: $1/TOL$.

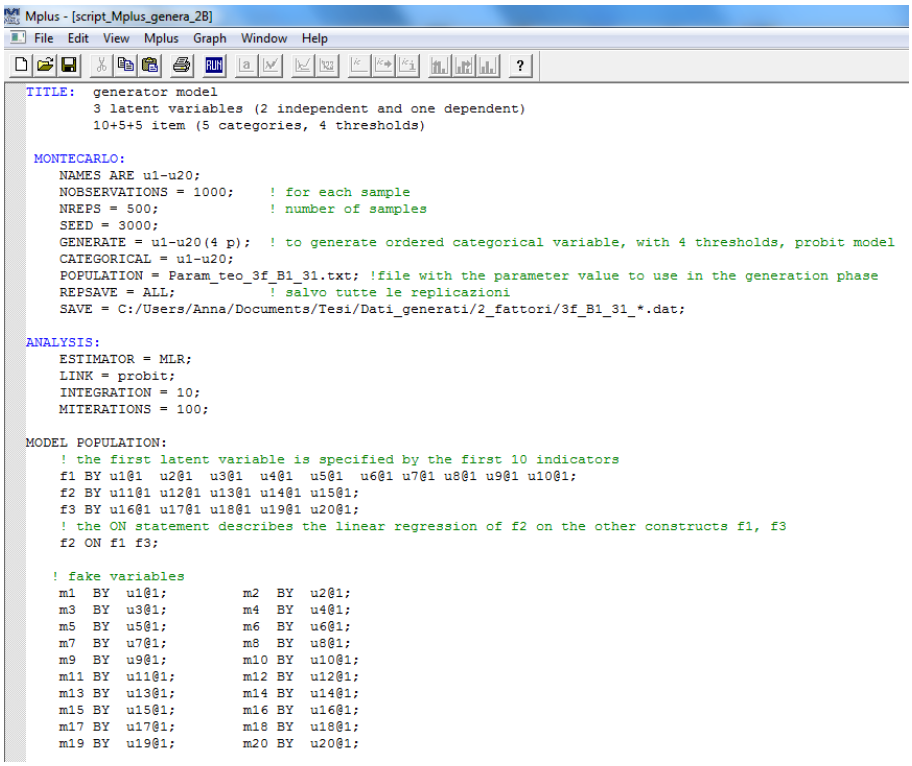
A final aspect to consider is the model identification. A model is said to be identified if it is impossible that a population variance-covariance matrix will be determined from two distinct sets of parameter values. A necessary condition for identification is that the number of distinct elements in the observed variance-covariance matrix is at least equal (better if it is greater) to the numbers of parameters to be estimated. This rule, which implies that the number of degrees of freedom be nonnegative, is easy to check, but unfortunately it is not a sufficient condition for identification. Even if a model is identified theoretically, there might be empirical identification problems. This happens when the expression of a parameter in terms of observed variances and covariances involves a denominator that is zero or close to zero (Kenny, 1979).

Appendix B

Script

B.1 M*plus* Script for the Two-step Procedure

B.1.1 M*plus* Script to Generate Data for the Second Scenario, Case B



```
    ! fixed at 0 the mean of the latent variables
    [f1@0];  [f2@0];  [f3@0];

    ! free the variance of the latent variables which is fixed at one as the default
    f1; f2; f3;

    ! fixed at 0 the correlation between the two independent latent variables
    f1 WITH f3@0;

    ! fixed at 0 the variance of the fake variable (as default, it will be 1)
    m1@0;  m2@0;  m3@0;  m4@0;  m5@0;  m6@0; m7@0; m8@0; m9@0; m10@0;
    m11@0; m12@0; m13@0; m14@0; m15@0;

    ! free the means of the fake variables which is fixed at 0 as the default
    [m1](c1);   [m2](c2);   [m3](c3);   [m4](c4);   [m5](c5);
    [m6](c6);   [m7](c7);   [m8](c8);   [m9](c9);   [m10](c10);
    [m11](c11); [m12](c12); [m13](c13); [m14](c14); [m15](c15);
    [m16](c16); [m17](c17); [m18](c18); [m19](c19); [m20](c20);

    ! to recreate a RSM, fix the threshold of item belonging to the first construct to be equal
    [u1$1 u2$1 u3$1 u4$1 u5$1 u6$1 u7$1 u8$1 u9$1 u10$1] (d1);
    [u1$2 u2$2 u3$2 u4$2 u5$2 u6$2 u7$2 u8$2 u9$2 u10$2] (d2);
    [u1$3 u2$3 u3$3 u4$3 u5$3 u6$3 u7$3 u8$3 u9$3 u10$3] (d3);
    [u1$4 u2$4 u3$4 u4$4 u5$4 u6$4 u7$4 u8$4 u9$4 u10$4] (d4);

    [u11$1 u12$1 u13$1 u14$1 u15$1] (d5);
    [u11$2 u12$2 u13$2 u14$2 u15$2] (d6);
    [u11$3 u12$3 u13$3 u14$3 u15$3] (d7);
    [u11$4 u12$4 u13$4 u14$4 u15$4] (d8);

    [u16$1 u17$1 u18$1 u19$1 u20$1] (d9);
    [u16$2 u17$2 u18$2 u19$2 u20$2] (d10);
    [u16$3 u17$3 u18$3 u19$3 u20$3] (d11);
    [u16$4 u17$4 u18$4 u19$4 u20$4] (d12);

   ! fixed at 0 the covariance between the fake variables
   m20  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0
             m14@0 m15@0 m16@0 m17@0 m18@0 m19@0;
   m19  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0
             m14@0 m15@0 m16@0 m17@0 m18@0;
   m18  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0
             m14@0 m15@0 m16@0 m17@0;
   m17  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0
             m14@0 m15@0 m16@0;
   m16  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0
             m14@0 m15@0;
   m15  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0
             m14@0 ;
   m14  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0 ;
   m13  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 ;
   m12  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 ;
   m11  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 ;
   m10  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 ;
   m9   WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 ;
   m8   WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 ;
   m7   WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 ;
   m6   WITH m1@0 m2@0 m3@0 m4@0 m5@0 ;
   m5   WITH m1@0 m2@0 m3@0 m4@0 ;
   m4   WITH m1@0 m2@0 m3@0 ;
   m3   WITH m1@0 m2@0 ;
   m2   WITH m1@0 ;

   ! fixed at 0 the covariance between the fake variables and the latent variables
   m1-m20 WITH f1@0 f2@0 f3@0;


MODEL CONSTRAINT:
  ! fixed the threshold of the first block of item to sum 0
  c1=-(c2 +c3 +c4 +c5 +c6 +c7 +c8 +c9 +c10);
  ! fixed the item difficulty of the first block of item to sum 0
  d4=-(d1 +d2 +d3);

  c11=-(c12 +c13 +c14 +c15);
  d8=-(d5 +d6 +d7);

  c16=-(c17 +c18 +c19 +c20);
  d12=-(d9 +d10 +d11);
```

B.1.2 M*plus* SEMirt Script to Analyze Data for the First Scenario, Case B

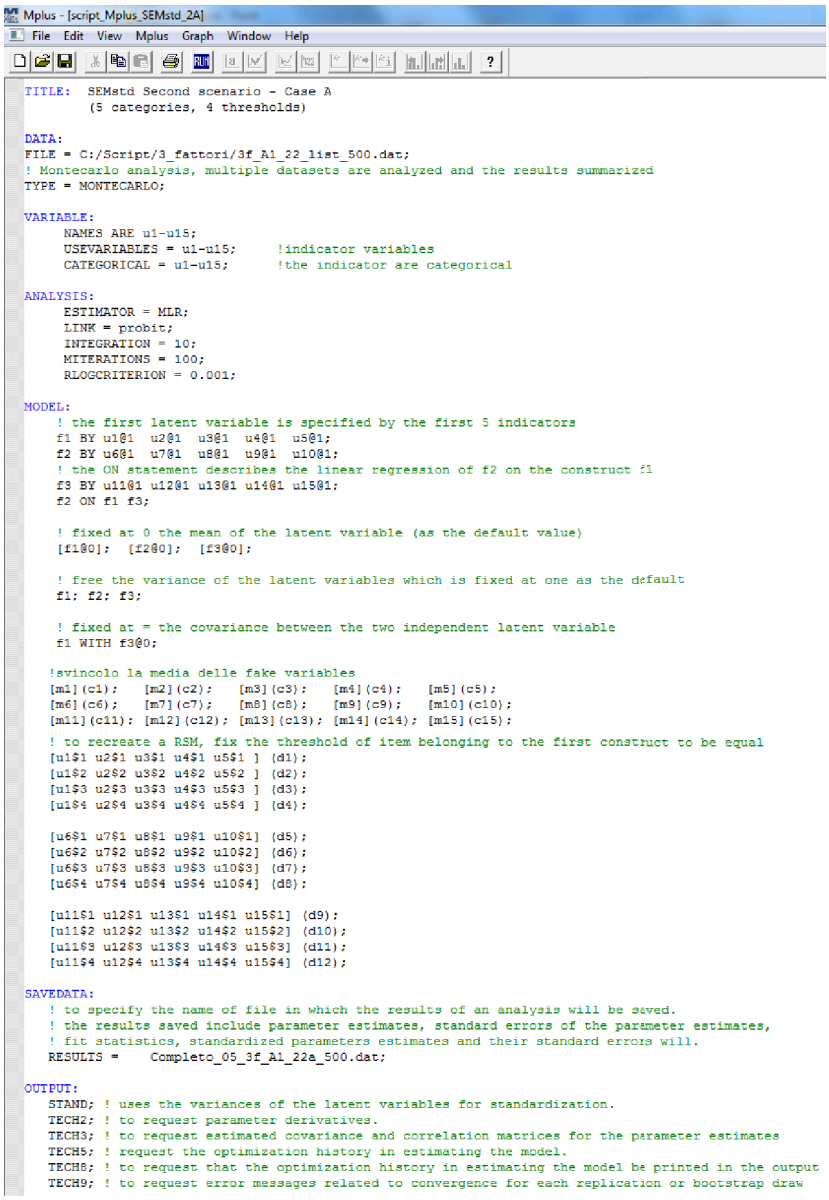
Mplus - [script_Mplus_SEMirt_1B]

File Edit View Mplus Graph Window Help

```
TITLE:  SEMirt First scenario - Case B
        (5 categories, 4 thresholds)

DATA:
FILE = C:/Dati_generati/2_fattori_050/b1_6.dat;
! Montecarlo analysis, multiple datasets are analyzed, results summarized using regular Mplus analysis
TYPE = MONTECARLO;

VARIABLE:
     NAMES ARE u1-u15;
     USEVARIABLES = u1-u15;      !indicator variables
     CATEGORICAL = u1-u15;       !the indicator are categorical

ANALYSIS:
     !MLR = ML parameter estimates with standard errors and a chi-square test statistic
     ESTIMATOR = MLR;
     LINK = probit;
     ! to specify number of integration points to be used in the computation
     INTEGRATION = 10;
     ! to specify the number of iterations allowed for the EM algorithm
     MITERATIONS = 100;
     ! to specify the relative observed data loglikelihood change convergence criterion for EM algorithm
     RLOGCRITERION = 0.001;

MODEL:
    ! the first latent variable is measured by the first 10 indicators
    f1 BY u1@1  u2@1  u3@1  u4@1  u5@1  u6@1 u7@1 u8@1 u9@1 u10@1;
    f2 BY u11@1 u12@1 u13@1 u14@1 u15@1;
    ! the ON statement describes the linear regression of f2 on the other construct f1
    f2 ON f1;

    ! free the variance of the latent variables which is fixed at one as the default
    f1; f2;

    ! fixed at 0 the variance of the fake variable (as default, it will be 1)
    m1@0;  m2@0;  m3@0;  m4@0;  m5@0;  m6@0; m7@0; m8@0; m9@0; m10@0;
    m11@0; m12@0; m13@0; m14@0; m15@0;

    !fake variables
    m1  BY  u1@1;          m2  BY  u2@1;
    m3  BY  u3@1;          m4  BY  u4@1;
    m5  BY  u5@1;          m6  BY  u6@1;
    m7  BY  u7@1;          m8  BY  u8@1;
    m9  BY  u9@1;          m10 BY u10@1;
    m11 BY  u11@1;        m12  BY  u12@1;
    m13 BY  u13@1;        m14  BY  u14@1;
    m15 BY  u15@1;

   !free the means of the fake variables which is fixed at 0 as the default
   [m1](c1);   [m2](c2);   [m3](c3);   [m4](c4);   [m5](c5);
   [m6](c6);   [m7](c7);   [m8](c8);   [m9](c9);   [m10](c10);
   [m11](c11); [m12](c12); [m13](c13); [m14](c14); [m15](c15);

   ! fixed at 0 the mean of the latent variable (as the default value)
   [f1-f2@0];
```

```
    ! to recreate a RSM, fix the threshold of item belonging to the first construct to be equal
    [u1$1 u2$1 u3$1 u4$1 u5$1 u6$1 u7$1 u8$1 u9$1 u10$1] (d1);
    [u1$2 u2$2 u3$2 u4$2 u5$2 u6$2 u7$2 u8$2 u9$2 u10$2] (d2);
    [u1$3 u2$3 u3$3 u4$3 u5$3 u6$3 u7$3 u8$3 u9$3 u10$3] (d3);
    [u1$4 u2$4 u3$4 u4$4 u5$4 u6$4 u7$4 u8$4 u9$4 u10$4] (d4);

    ! to recreate a RSM, fix the threshold of item belonging to the first construct to be equal
    [u11$1 u12$1 u13$1 u14$1 u15$1] (d5);
    [u11$2 u12$2 u13$2 u14$2 u15$2] (d6);
    [u11$3 u12$3 u13$3 u14$3 u15$3] (d7);
    [u11$4 u12$4 u13$4 u14$4 u15$4] (d8);

    ! fixed at 0 the covariance between the fake variables
    m15  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0 m14@0 ;
    m14  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 m13@0 ;
    m13  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 m12@0 ;
    m12  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 m11@0 ;
    m11  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 m10@0 ;
    m10  WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 m9@0 ;
    m9   WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 m8@0 ;
    m8   WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 m7@0 ;
    m7   WITH m1@0 m2@0 m3@0 m4@0 m5@0 m6@0 ;
    m6   WITH m1@0 m2@0 m3@0 m4@0 m5@0 ;
    m5   WITH m1@0 m2@0 m3@0 m4@0 ;
    m4   WITH m1@0 m2@0 m3@0 ;
    m3   WITH m1@0 m2@0 ;
    m2   WITH m1@0 ;

    m1-m15 WITH f1@0 f2@0;


MODEL CONSTRAINT:
    ! fixed the threshold of the first block of item to sum 0
    c1=-(c2 +c3 +c4 +c5 +c6 +c7 +c8 +c9 +c10);
    ! fixed the item difficulty of the first block of item to sum 0
    d4=-(d1 +d2 +d3);

    c11=-(c12 +c13 +c14 +c15);
    d8=-(d5 +d6 +d7);

OUTPUT:
    !uses the variances of the latent variables for standardization.
    STAND;
    TECH2; ! to request parameter derivatives.
    TECH3; ! to request estimated covariance and correlation matrices for the parameter estimates
    TECH5; ! request the optimization history in estimating the model.
    TECH8; ! to request that the optimization history in estimating the model be printed in the output
    TECH9; ! to request error messages related to convergence for each replication or bootstrap draw
```

B.1.3 M*plus* SEMstd Script to Analyze Data for the Second Scenario, Case A



B.2 *R* Script for the Two-step Procedure

B.2.1 *R* RSM-LRMme Script

```
#####   SCRIPT for RSM-LRMme
#####  first scenario - case B

# to empty the memory
rm(list=ls())

library(eRm)     # load package for the RSM
library(psy)     # load package for calculate the Cronbach's coefficient alpha

start<-Sys.time()

reliability <- "_050"
inizio <- "050_"

# directory to read and save data
dir_dati <- paste("C:/Dati_generati/2_fattori", reliability, "/", sep="")
dir_FS   <- paste("C:/Dati_generati/Dati_FS", reliability, "/", sep="")
dir_RAM  <- paste("C:/Dati_generati/Dati_RAM", reliability, "/", sep="")
dir_result <- paste("C:/Dati_generati/Analisi_R/IRT_Fox",reliability,"/",sep="")

n1 <- 10  # nr item for the first construct
n2 <- 5   # nr item for the second construct
cat1 <- 5 # nr categories for the first construct
cat2 <- 5 # nr categories for the first construct

# indentify different parameter config. (CP) (case B, psi2=0.1,0.3,0.5,0.7,0.9)
gruppi <- c("B1",   "B3",   "B5",   "B7",   "B9")
L <- length(gruppi)

# matrix to save means and std deviation of beta coefficient for each CP
mat_param_riass <- matrix(0,L,9)
colnames(mat_param_riass) <- c("gruppo","n°medio_oss","beta_corr","var_f1",
                               "var_f2","var_e1","var_e2","cronbach_1","cronbach_2")

################################################################################
################    loop for each parameter configuration     ##################
################################################################################

for(m in 1:L){
```

```
gruppo <- gruppi[m]

# name of file that contains the list of the file to analyze (CP)
nome_lista_leggere <- paste(dir_dati, inizio, gruppo, "_list.dat", sep="")

nome_lista_salvati <- paste(dir_FS, inizio, "list_FS_", gruppo, ".dat",sep="")
nome_lista_RAM   <- paste(dir_RAM, inizio, "list_RAM_", gruppo, ".dat",sep="")
nome_alpha     <- paste(dir_risultati, inizio, "alpha_",gruppo, ".dat",sep="")

#read the single file to analyze
lista_file <- read.table(nome_lista_leggere, as.is=1)
n_file <- dim(lista_file)[1]

# matrix to save the Cronbach's alpha value for the 2 measures
mat_alpha <- matrix(0,n_file,2)

# matrix to save item difficulty and thresholds
mat_soglie_1 <- matrix(0,n_file,(n1+(cat1-1)))
mat_soglie_2 <- matrix(0,n_file,(n2+(cat1-1)))

# vector of filename of factor scores
file_salvati <- c()

#vector of filename of RAM matrix
file_RAM <- c()

# general RAM matrix to use in the SEM analysis
input_RAM <- matrix("NA",7,3)
input_RAM[1,1] <- "F1 -> y1, NA,"
input_RAM[1,2] <- 1
input_RAM[1,3] <- ","

input_RAM[2,1] <- "F2 -> y2, NA,"
input_RAM[2,2] <- 1
input_RAM[2,3] <- ","

input_RAM[3,1] <- "F1 -> F2, beta,"
input_RAM[3,2] <- "NA"
input_RAM[3,3] <- ","
input_RAM[4,1] <- "F1 <-> F1, var_f1,"
input_RAM[4,2] <- 1
input_RAM[4,3] <- ","

input_RAM[5,1] <- "F2 <-> F2, var_f2,"
input_RAM[5,2] <- 1
input_RAM[5,3] <- ","

input_RAM[6,1] <- "y1 <-> y1, NA,"
input_RAM[6,2] <- 1
input_RAM[6,3] <- ","

input_RAM[7,1] <- "y2 <-> y2, NA,"
input_RAM[7,2] <- 1
input_RAM[7,3] <- ","

n_oss_file <- 0
```

```
##############################################################################
################      loop IRM for each file in the CP      ##################
##############################################################################

# j is the number of sample for each CP (in our simulation is 500)

for(j in 1:n_file){

  # read data
  nome_file <- paste(dir_dati, lista_file[j,1],sep="")
  dati1 <- data.frame(read.table(nome_file))
  n <- dim(dati1)[1]

  # we delete from the dataset the subect with extreme response pattern for
  # the 1° or 2° construct because RSM does not consider these subjects
  max_risposta <- max(dati1)
  min_risposta <- min(dati1)

  a <-which(rowSums(dati1[,1:n1])==0 | rowSums(dati1[,1:n1])==max_risposta*n1)
  b <-which(rowSums(dati1[,(n1+1):(n1+n2)])==0|rowSums(dati1[,(n1+1):(n1+n2)])
            ==max_risposta*n2)
  ab <- c(a,b)
  if (length(ab)>0){
    dati1<-dati1[-ab,]
  }

  n_reale <- dim(dati1)[1]
  n_oss_file <- n_oss_file+n_reale

  ### Estimate of Y1 - first Rasch measure (item from 1 to n1)
  rsm_y1 <- RSM(dati1[,1:n1],sum0 = TRUE)
  soglie_1 <- thresholds(rsm_y1)$threshtable[[1]]     # item thresholds
  for(k in 1:n1){                # item difficulty
    mat_soglie_1[j,k] <- soglie_1 [k,1]
  }
  for(k in (n1+1):(n1+(cat1-1))){
    mat_soglie_1[j,k] <- soglie_1 [1,(k-n1+1)]-soglie_1[1,1]
  }
  person_y1   <- person.parameter(rsm_y1)
  p1    <- (person_y1$thetapar)$NAgroup1      # first Rasch meaasure Y1
  var_p1      <- (var(p1)*(n_oss_file-1))/n_oss_file
  p1_std  <- (p1-mean(p1))/sqrt(var_p1)       # standardize Y1
  var_p1_std  <- (var(p1_std)*(n_oss_file-1))/n_oss_file

         # std error of person parameter (used to estimate measurement error)
  se_p1   <- (person_y1$se.theta)$NAgroup1
  per_par_y1  <- cbind(p1_std,se_p1)


  ### Estimate of Y2 - seconf Rasch measure (item from n1+1 to n2)
  rsm_y2  <- RSM(dati1[,(n1+1):(n1+n2)])
  soglie_2 <- thresholds(rsm_y2)$threshtable[[1]]
  for(k in 1:n2){
    mat_soglie_2[j,k] <- soglie_2[k,1]
  }
  for(k in (n2+1):(n2+(cat2-1))){
    mat_soglie_2[j,k] <- soglie_2[1,(k-n2+1)]-soglie_2[1,1]
  }
```

```
  person_y2   <- person.parameter(rsm_y2)
  p2    <- (person_y2$thetapar)$NAgroup1
  var_p2      <- (var(p2)*(n_oss_file-1))/n_oss_file
  p2_std  <- (p2-mean(p2))/sqrt(var_p2)
  var_p2_std  <- (var(p2_std)*(n_oss_file-1))/n_oss_file
  se_p2   <- (person_y2$se.theta)$NAgroup1
  per_par_y2  <- cbind(p2_std,se_p2)


  per_y1_MSE  <- (sum(se_p1^2)/n_reale)/var_p1    # MSME(Y1)
  per_y1_SA <- var_p1_std-(per_y1_MSE)      # variance of theta1
  per_y1_Rel  <- (per_y1_SA)/var_p1_std     # reliability of Y1

  per_y2_MSE  <- (sum(se_p2^2)/n_reale)/var_p2    # MSME(Y2)
  per_y2_SA <- var_p2_std-(per_y2_MSE)      # variance of theta2
  per_y2_Rel  <- (per_y2_SA)/var_p2_std     # reliability of Y2


  #insert the measurement error variances into the RAM matrix for the 2° step
  input_RAM[6,2] <- per_y1_MSE
  input_RAM[7,2] <- per_y2_MSE

  # Rasch measures and std errors matrix
  person_par <- cbind(per_par_y1,per_par_y2)

  # Cronbach's alpha of Y1 and Y2
  mat_alpha[j,1]<-cronbach(dati1[,1:n1])$alpha
  mat_alpha[j,2]<-cronbach(dati1[,(n1+1):(n1+n2)])$alpha

  # save data
  nome_file_FS <- paste(inizio, gruppo,"_FS_",j,".dat",sep="")
  write.table(person_par, file=paste(dir_FS, nome_file_FS, sep=""), sep=" ",
              row.names=TRUE,col.names=FALSE)

  nome_file_RAM <- paste(inizio, gruppo, "_RAM_",j,".dat",sep="")
  write.table(input_RAM, file=paste(dir_RAM, nome_file_RAM, sep=""), sep=" ",
              row.names=FALSE,col.names=FALSE, quote=FALSE)

  file_salvati <- rbind(file_salvati, nome_file_FS)
  file_RAM <- rbind(file_RAM, nome_file_RAM)

}

############################## End IRT loop  ###################################


# save data summary

write.table(file_salvati, file=nome_lista_salvati, sep="",row.names=FALSE,
            col.names=FALSE, quote=FALSE)
write.table(file_RAM, file=nome_lista_RAM, sep="",row.names=FALSE,quote=FALSE)
write.table(mat_alpha, file=nome_alpha, sep="",row.names=FALSE,quote=FALSE)

mat_soglie <- cbind(mat_soglie_1,mat_soglie_2)
nome_file_mat_soglie <-paste(dir_risultati,"mat_soglie_",gruppo,".xls",sep="")
write.table(mat_soglie,file=nome_file_mat_soglie,row.names=FALSE,quote=FALSE)
```

```
##############################    SEM loop    ###################################


library(sem)       # package for SEM

nome_lista_leggere <- paste(dir_FS, inizio, "list_FS_", gruppo,".dat", sep="")

# matrice RAM senza varianze latente (da stimare)
nome_lista_RAM   <- paste(dir_RAM, inizio, "list_RAM_", gruppo,".dat", sep="")

# matrix of beta coefficient
mat_param <- matrix(0,1,8)
colnames(mat_param) <- c("n°file", "beta_corr", "var_f1", "var_f2", "var_e1",
                         "var_e2","cronbach_1","cronbach_2")

lista_file_FS <- read.table(nome_lista_leggere,as.is=1)
lista_file_RAM <- read.table(nome_lista_RAM,as.is=1)

n_file <- dim(lista_file_FS)[1]


###############################################################################
##################     loop SEM for each file in the CP    ####################
###############################################################################

for(j in 1:n_file){

  nome_file_FS <- paste(dir_FS, lista_file_FS[j,1], sep="")
  nome_file_RAM <- paste(dir_RAM, lista_file_RAM[j,1], sep="")

  # read data, select Factor scores and covert into dataframe
  dati <- read.table(nome_file_FS, as.is=1)
  dati_FS <- as.data.frame(cbind(dati[,2],dati[,4]))
  colnames(dati_FS) <- c("y1","y2")

  N <- nrow(dati_FS)-1

  # Variance-covariance matrix
  Cov <- with(dati_FS, cov(cbind(dati_FS[1],dati_FS[2])))

  # load RAM matrix (that specify the structural and the measurement model)
  RAM <- specify.model(nome_file_RAM)

  # SEM estimation
  sem <- sem(RAM, Cov, N, raw=FALSE, debug=FALSE, warn=TRUE)

  parametri <- sem$coeff      # beta coefficient, variance of theta1 & theta2
  beta <- parametri[1]
  var_f1 <- parametri[2]
  var_f2 <- parametri[3]+beta^2*var_f1
```

```
    # loop to obtain and to verify the consistency of the parameter estimates
    valore_iniziale <- 1
    while(var_f1<0 | var_f2<0){
      valore_iniziale <- valore_iniziale-0.1
      RAM[4,3] <- valore_iniziale
      RAM[5,3] <- valore_iniziale
      sem <- sem(RAM, Cov, N, raw=FALSE, debug=FALSE, warn=TRUE)
      parametri   <- sem$coeff    # coeff beta, varianza di f1 e varianza di f2
      beta <- parametri[1]
      var_f1 <- parametri[2]
      var_f2 <- parametri[3]+beta^2*var_f1

    }

    e1 <- as.numeric(RAM[6,3])      # (1-RPRI) of Y1
    e2 <- as.numeric(RAM[7,3])      # (1-RPRI) of Y2

    mat_param <- rbind(mat_param, cbind(j, beta, var_f1, var_f2, e1, e2,
                       mat_alpha[j,1], mat_alpha[j,2]))
  }
  ############################### End SEM loop  ################################

  # save coefficient matrix
  nome_file_param <- paste(dir_risultati, "Parametri_", gruppo, ".xls", sep="")
  mat_param <- mat_param[-1,]
  write.table(mat_param, file=nome_file_param, row.names=FALSE,quote=FALSE)

  mat_param_riass[m,] <- cbind(gruppo, (n_oss_file/n_file),mean(mat_param[,2]),
                 mean(mat_param[,3]), mean(mat_param[,4]),mean(mat_param[,5]),
                 mean(mat_param[,6]), mean(mat_param[,7]),mean(mat_param[,8]))

}

##########################   end loop  for each CP   ########################

nome_file_risultati <- paste(dir_risultati, reliability,"ris_simul.xls",sep="")
write.table(mat_param_riass, file=nome_file_risultati, row.names=FALSE,
            col.names=TRUE, quote=FALSE)

end<-Sys.time()

# to see the elaboration time
start
end
```

B.2.2 *R* RSM-LRM Script, only Second Step

```
#####   SCRIPT for LRM
#####   second scenario - case A and B

# to empty the memory
rm(list=ls())
library(eRm)

start <- Sys.time()
reliability <- "_050"
inizio <- "050_"

dir_regr <- "C:/Dati_generati/Analisi_R/Regressione/"
dir_FS      <-  paste("C:/Dati_generati/Dati_FS_3f", reliability, "/", sep="")

# all the 25 CP of the second scenario
gruppi <- c("3f_A1_13", "3f_A1_22", "3f_A3_13","3f_A3_22","3f_A5_13",
    "3f_A5_22", "3f_A7_13", "3f_A7_22", "3f_A9_13", "3f_A9_22",
    "3f_B1_13", "3f_B1_22", "3f_B1_31", "3f_B3_13", "3f_B3_22",
    "3f_B3_31", "3f_B5_13", "3f_B5_22", "3f_B5_31", "3f_B7_13",
    "3f_B7_22", "3f_B7_31", "3f_B9_13", "3f_B9_22", "3f_B9_31")
L <- length(gruppi)

# matrix to save means and std deviation of beta coefficient for each CP
mat_param_riass <- matrix(0,L,6)
colnames(mat_param_riass) <- c("gruppo", "n°medio_oss", "beta13", "beta23",
                               "beta13_corr", "beta23_corr")

# matrices and vectors to save the parameters
coeff_beta <- matrix(0,1000,L*2)
coeff_beta_corr <- matrix(0,1000,L*2)
nomi_coeff_beta  <- c()
nomi_coeff_beta_corr <- c()
for (i in 1:L){
  nomi_coeff_beta      <- cbind(nomi_coeff_beta, paste(gruppi[i], "_beta13",
                          sep=""), paste(gruppi[i], "_beta23", sep=""))
  nomi_coeff_beta_corr <- cbind(nomi_coeff_beta_corr,paste(gruppi[i],"_beta13",
                          sep=""), paste(gruppi[i], "_beta23", sep=""))
}

colnames(coeff_beta) <- nomi_coeff_beta
colnames(coeff_beta_corr) <- nomi_coeff_beta_corr

##############################################################################
###################            loop foreach CP           #####################
##############################################################################

for(m in 1:L){

  gruppo <- gruppi[m]

  nome_lista_leggere <- paste(dir_FS, inizio,"list_FS_", gruppo, ".dat",sep="")
  lista_file_FS <- read.table(nome_lista_leggere,as.is=1)
  n_file <- dim(lista_file_FS)[1]

  n_oss_file <- 0
```

Acknowledgments

This work is the result of a long period of research and study, during which I have been supported, encouraged and spurred on by numerous people. First and foremost, I would like to thank my thesis advisor Professor Maurizio Carpita, his valuable knowledge in the field has allowed me to investigate many of the issues covered. He taught me how to approach solving problems with an open mind and out of sincere curiosity, and how to investigate questions carefully and with enthusiasm. I am extremely grateful for his careful reading of several versions of the thesis, his helpful advice, and for always "leaving his doors open".

I must also thank all the people of the statistical section of the Department of Quantitative Methods of the University of Brescia, for their support and their professional experience.

Finally I must thank my family and Claudio, who have always wished the best for me, patiently waiting for that I successfully conclude this work and I achieve this important goal.

Bibliography

Abreyava, J. and Hausman, J. (2004). Response error in a transformation model with an application to earnings equation estimation. *Econometrics Journal*, 7:366–388.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Wiley-Interscience.

Allen, M. J. and Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing Co., Monterry, CA.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112(4):545 – 557.

Anderson, J. C. and Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3):411–423.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43:561–573.

Andrich, D. (2005). The rasch model explained. In Alagumalai, S., Durtis, D. D., and Hungi, N., editors, *Applied Rasch Measurement: A book of exemplars*, pages 308–328. Springer-Kluwer.

Baker, F. B. (1985). *The Basics of Item Response Theory*. Heinemann, Portsmouth, NH.

Baker, F. B. and Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques*. CRC Press, New York, 2nd edition.

Bartholomew, D. (1996). *The Statistical Approach to Social Measurement*. Academic Press, San Diego.

Bentler, P. M. and Yuan, K.-H. (1999). On adding a mean structure to a covariance structure model.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, New York.

Bollen, K. A. and Curran, P. (2006). *Latent Curve Models: A Structural Equation Perspective*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Bond, T. G. and Fox, C. M. (2001). *Applying the Rasch model : fundamental measurement in the human sciences*. Lawrence Erlbaum Ass.; London.

Browne, M. W. (1974). Generalized least squares estimates in the analysis of covariance structures. *South African Statistical Journal*, 8:1–24.

Buonaccorsi, J. P. (1996). Measurement error in the response in the general linear model. *Journal of the Americal Statistical Association*, 91(434):633–642.

Buzas, J. S., Tosteson, T. D., and Stefanski, L. A. (2003). Measurement error. Technical Report 2544, Institute of Statistics Mimeo Series.

Cagnone, S., Mignani, S., and Moustaki, I. (2010). Latent variable models for ordinal data. In Monari, P., Bini, M., Piccolo, D., and Salmaso, L., editors, *Statistical Methods for the Evaluation of Educational Services and Quality of Products*, Contributions to Statistics, pages 17–28. Physica-Verlag HD.

Carpita, M. (2009). *La qualità del lavoro nelle cooperative sociali. Misure e modelli statistici*. FrancoAngeli, Milano.

Carroll, R., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A Modern Perspective*. Chapman & Hall/CRC, London, 2nd edition.

Carroll, R. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, 85(411):652–663.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and weighting in regression*. Chapman and Hall, London.

Cook, J. and Stefanski, L. (1994). Simulationextrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89:1314–1328.

Copas, J. (1988). Binary regression models for contamined data (with discussion. *Journal of the Royal Statistical Society, Series B*, 50(2):225–265.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

De Leeuw, J. and Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11:57–85.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3):292–307.

Devanarayan, V. and Stefanski, L. A. (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics & Probability Letters*, 59(3):219 – 225.

DeVellis, R. (1991). *Scale Development: Theory and Application*. Sage Publications, Newbury Park.

Emons, W., Sijtsma, K., and Meijer, R. (2005). Global, local and graphical person-fit analysis using person response functions. *Psychological Methods*, 10(1).

Faliva, M. (1987). *Econometria: principi e metodi*. UTET, Torino.

Ferrando, P. J. and Chico, E. (2007). The external validity of scores based on the two-parameter logistic model: Some comparisons between irt and ctt. *Psicológica*, 28:237–257.

Fischer, G. H. and Ponocny, I. (1994). An extension of the partial credit model with an application to the measurement of change. *Psychometrika*, 59:177–192.

Fisher Jr, W. P. (1992). Objectivity in measurement: A philosophical history of raschs separability theorem. In Wilson, M., editor, *Objective Measurement: Theory into Practice, Volume 1*, pages 29–60. Ablex, Norwood, NJ.

Fox, J. (2006). Teacher's corner: Structural equation modeling with the sem package in r. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(3):465–486.

Fraser, G. E. and Stram, D. O. (2001). Regression calibration in studies with correlated variables measured with error. *American Journal of Epidemiology*, 154(9):836–844.

Fuller, W. (1987). *Measurement error models*. John Wiley & Sons, New York.

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. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31(1):4–19.

Goldberger, A. S. and Duncan, O. D. (1973). *Structural equation models in the social sciences*. Seminar Press, New York.

Goldstein, H. (1989). Restricted unbiased iterative generalized least-squares estimation. *Biometrika*, 76(3):622–623.

Goldstein, H. (1995). *Multilevel Statistical Models*. Kendall's Library of Statiscs 3. Edward Arnold, London, 2nd edition.

Goldstein, H. and Rasbash, J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares. *Computational Statistics and Data Analysis*, 13:63–71.

Gulliksen, H. (1945). The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 10(2):79–91.

Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Press, Newbury Park, CA.

Hardin, J. W., Schmiediche, H., and Carroll, R. J. (2003). The simulation extrapolation method for fitting generalized linear models with additive measurement error. *Stata Journal*, 3(4):373–385.

Harris, D. (1989). Comparison of 1-, 2-, and 3-paramter irt models. *Educational Measurement: Issues and Practice*, 8:35–41.

Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2):139–164.

Hatzinger, R. (2008). A glm framework for item response theory models. Technical Report 66, Department of Statistics and Mathematics.

Hox, J. (2002). *Multilevel Analysis. Techniques and Applications*. Lawrence Erlbaum Associates, Mawah NJ.

Huber, P., Ronchetti, E., and Victoria-Feser, M.-P. (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66:893–908.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Jones, M. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91:222–230.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In Goldberger, A. S. and Duncan, O. D., editors, *Structural Equation Models in the Social Sciences*, pages 85–112. Academic Press, New York: Academic Press.

Jöreskog, K. G. and Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36:347–387.

Jöreskog, K. G. and Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Abt Books, Cambridge, MA.

Keesling, J. W. (1972). *Maximum Likelihood Approaches to Causal Analysis*. PhD thesis, University of Chicago, Chicago.

Kenny, D. A. (1979). *Correlation and causality*. Wiley, New York.

Küchenhoff, H., Mwalili, S., and Lesaffre, E. (2006). A general method for dealing with misclassification in regression: the misclassification simex. *Biometrics*, 62:85–96.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago Press, Chicago.

Linacre, J. M. (1997). Kr-20 / cronbach alpha or rasch reliability: Which tells the "truth"? *Rasch Measurement Transactions*, 11(3):580–1.

Linacre, J. M. (2003). *Users guide and program manual to WINSTEPS: Rasch model computer programs*. MESA Press, Chicago.

Linacre, J. M. and Wright, B. D. (1995). *A users guide to BIGSTEPS: Rasch model computer programs*. MESA Press, Chicago.

Lindley, D. and Novick, M. (1981). The role of exchangeability in inference. *Annals of Statistics*, 9:45–58.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(1):1–41.

Liou, M. (1994). More on the computation of higher-order derivatives of the elementary symmetric functions in the rasch model. *Applied Psychological Measurement*, 18:53–62.

Littell, R., Milliken, G., Stroup, W., and Wolfinger, R. (1996). *SAS System for Mixed Models*. SAS Institute Inc, Cary, NC.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data, Second Edition*. Wiley-Interscience, New York, 2nd edition.

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, 51(493-504).

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74:817–827.

Longford, N. T. (1994). *Random Coefficient Models*. Oxford University Press, Oxford.

Lord, F. (1952). *A theory of test scores*. Psychometric Monographs, Chicago.

Lord, F. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum, Hillsdale, NJ.

MacCallum, R. and Austin, J. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51:201–226.

Mair, P. and Hatzinger, R. (2007a). Extended rasch modeling: The erm package for the application of irt models in r. *Journal of Statistical Software*, 20(9):1–20.

Mair, P. and Hatzinger, R. (2007b). Cml based estimation of extended rasch models with the erm package in r. *Psychology Science*, 49:26–43.

Martin, J. A. (1987). Structural equation modeling: A guide for the perplexed. *Child Development*, 58:33–37.

Masters, G. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47:149–174.

McArdle, J. J. (1980). Causal modeling applied to psychonomic systems simulation. *Behavior Research Methods and Instrumentation*, 12:193–209.

McArdle, J. J. and McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *The British journal of mathematical and statistical psychology*, 37(2):234–251.

McDonald, R. P. and Hartmann, W. M. (1992). A procedure for obtaining initial values of parameters in the ram model. *Multivariate Behavioral Research*, 27:57–76.

Meyer, P. J. (2000). *Reliability*. Oxford University Press, Oxford.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43(4):551–560.

Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association*, 27:807–811.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22(1-2):43–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1):115–132.

Muthén, B. and Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, 46(4):407–419.

Muthén, L. K. and Muthén, B. O. (2007). *Mplus Statistical Analysis With Latent Variables. Users Guide*. Muthén & Muthén, Los Angeles.

Nering, M. and Ostini, R. (2010). *Handbook of Polytomous Item Response Theory Models*. Routledge Academic.

Neuhaus, J. M. (2002). Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics*, 58:675–683.

Novick, M. and Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32:1–13.

Nunnally, J. C. and Bernstein, I. H. (1994). *Psychometric Theory*. McGraw-Hill, New York, 3rd edition.

Paulino, C. D., Soares, P., and Neuhaus, J. (2003). Binomial regression with misclassification. *Biometrics*, 59:670–675.

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge University Press, New York.

Pfanzagl, J. (1994). On item parameter estimation in certain latent trait models. In Fischer, G. and Laming, D., editors, *Contributions to Mathematical Psychology, Psychometrics, and Methodology*, pages 249–263. Springer, New York.

Prescott, G. J. and Garthwaite, P. H. (2002). A simple bayesian analysis of misclassified binary data with a validation study. *Biometrics*, 58:454–458.

Ramalho, E. A. (2002). Regression models for choice-based samples with misclassification in the response variable. *Journal of Econometrics*, 106:171–201.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the IV. Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV*, pages 321–333. University of California Press, Berkeley.

Raudenbush, S. and Bryk, A. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59:1–17.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *american sociological review*, 15:351–357.

Roju, N. S., Van der Linden, W. J., and Fleer, P. F. (1995). Irt-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19:353–368.

Rosner, B., Spiegelman, D., and Willett, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American journal of epidemiology*, 132(4):734–745.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34:100–114.

Schaubroeck, J. (1990). Investigating reciprocal causation in organizational behavior research. *Journal of Organizational Behavior*, 11:17–28.

Schumacker, R. E. and Smith Jr., E. V. (2007). Reliability - a rasch perspective. *Educational and Psychological Measurement*, 67(3):394–409.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC.

Spearman, C. (1904). "general intelligence," objectively determined and measured. *American Journal of Psychology*, 15:201–293.

Stefanski, L. and Carrol, R. (1985). Covariate measurement error in logistic regression. *Annals of statistics*, 12:1335–1351.

Tarkkonen, L. and Vehkalahti, K. (2005). Measurement errors in multivariate measurement scales. *J. Multivar. Anal.*, 96(1):172–189.

Teel, J. E., Bearden, W. O., and Sharma, S. (1986). Interpreting lisrel estimates of explained variance in nonrecursive structural equation models. *Journal of Marketing Research*, 23(2):164–168.

Thompson, B., editor (2003). *Score Reliability. Contemporary Thinking on Reliability issues*. Sage Publications, London.

Van der Linden, W. and Hambleton, R. (1997). *Handbook of Modern Item Response Theory*. Springer, New York.

Vogt, W. P. (1993). *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences*. SAGE Publications, Newbury Park, CA.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In Goldberger, A. S. and Duncan, O. D., editors, *Structural Equation Models in the Social Sciences*. Academic Press, New York.

Woodhouse, G., Yang, M., Goldstein, H., and Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society. Series A*, 159(2):202–212.

Wright, B. D. (1999). Model selection: Rating scale or partial credit? *Rasch Measurement Transactions*, 12(3):641–642.

Wright, B. D. and Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement Series*. Chicago: MESA.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5:161–215.

Zarate, L., Nogueira, B., Santos, T., and Song, M. (2007). Techniques for missing value recovering in imbalanced databases: Application in a marketing database with massive missing data. In *IEEE International Conference on Systems and Man and Cybernetics*, pages 2658–2664.

Zumbo, B. (2007). Three generations of differential item functioning (dif) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4:233–233.