Identification of Parametric Rasch-Type Models

ERNESTO SAN MARTÍN & JEAN MARIE-ROLIN

December 30, 2010
Identification of Parametric Rasch-type Models

ERNESTO SAN MARTÍN\textsuperscript{a,b} AND JEAN-MARIE ROLIN\textsuperscript{c}

\textsuperscript{a}Department of Statistics, Pontificia Universidad Católica de Chile, Chile.
\textsuperscript{b}Measurement Center MIDE UC, Pontificia Universidad Católica de Chile, Chile.
\textsuperscript{c}Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Belgium.

December 30, 2010

Abstract

Recently, San Martín et al. (2010) have shown that a Bayesian nonparametric generalization of a Rasch model has not empirical meaning in the sense that the distribution generating the person-specific random effects is unidentified. This result is an invitation to reconsider parametric specifications of Rasch models, particularly their identifiability. This paper provides an explicit proof of the identifiability of parametric generalizations of Rasch models. Furthermore, these identification results are extended to explanatory IRT-models at the item side as well as at the person side.

Keywords: Location-scale distributions; Generalized Linear Mixed Model; Identified Parameter; Parameter of Interest; Explanatory IRT-Models.

1 Introduction

Generalized Linear Mixed Models (GLMM) are widely used in many fields, notably in biometrics and psychometrics. The comprehensive books by Verbeke and Molenberghs (2009) and De Boeck and Wilson (2004) are relevant references. This class of models is formulated in a hierarchical way as follows: let $Y_{pi}$ denote the $i$-th measurement available for the $p$-th person, and let $Y_p$ denote the corresponding vector of all measurements. It is assumed that

\[ (Y_p \mid \theta_p, \beta) \sim F_p^{\beta} \]  \hspace{1cm} (1.1)

where $F_p$ is a pre-specified distribution, parametrized through a vector $\beta$ of unknown parameters, common to all persons. Further, $\theta_p$ is a unidimensional variable of person-specific characteristics, called random effect, assumed to follow a so-called mixing distribution $G$ which is typically considered to be a member of a parametrized family, namely

\[ (\theta_p \mid \sigma) \overset{iid}{\sim} G^\sigma. \]  \hspace{1cm} (1.2)
In psychometrics, the Rasch model is an important example of this type of models (Ghosh et al., 2000; De Boeck and Wilson, 2004). Defining $\beta = (\beta_1, \ldots, \beta_I)$, the conditional model (1.1) is specified as

$$(Y_{pi} | \theta_p, \beta_i) \sim \text{Bern}(\Psi(\theta_p - \beta_i)), \quad p = 1, \ldots, N; \ i = 1, \ldots, I,$$  

(1.3)

where, for each person $p$, $\{Y_{pi} : i = 1, \ldots, I\}$ are mutually independent conditionally on $(\theta_i, \beta)$ and $\Psi(x) = \exp(x)/(\exp(x) + 1)$. In the psychometric literature, the conditional independence of the $Y_{pi}$’s given $(\theta_i, \beta)$ is known as the axiom of local independence (Lazarsfeld, 1954) and it is related with the hypothesis of common cause already used by Laplace (1774) in the context of the rules of succession. For the marginal model (1.2), it is typically assumed that the random effects $\theta_p$’s are iid $\mathcal{N}(0, \sigma^2)$; these random effects reflect the between-person heterogeneity in the population with respect to the distribution of $Y_p$ (Molenbergs and Verbeke, 2004, p.120).

This hierarchical structure is intended to model the measurements $Y_p$ through a statistical model induced after integrating out the random effects, namely

$$P(Y_p = y | \beta, \sigma) = \int \prod_{i=1}^{I} \left[ \Psi(\sigma \theta - \beta_i)^{y_i} [1 - \Psi(\sigma \theta - \beta_i)]^{1-y_i} \phi(\theta) \right] d\theta,$$

(1.4)

where $y \in \{0, 1\}^I$ and $\phi(\cdot)$ corresponds to the density function of a standard normal distribution.

In such a case, $(\beta, \sigma)$ are estimated using available software packages as SAS, Stata, R or BUGS. However, in applications there are almost inevitable concerns about the lack of robustness of resulting inferences with respect to assumed forms of distributional components (Woods and Thissen 2006; Agresti et al. 2004; Heagerty and Kurland 2001; Verbeke and Lesaffre 1996). Therefore, the assumption of a specific parametric distribution may be too restrictive to represent the actual between-person variation. These concerns lead to consider the items parameters $\beta$ and a general probability distribution $G$ generating the random effects as parameters of interest, and have motivated a rich literature on classical (Woods and Thissen, 2006; Woods, 2006, 2008) and Bayesian nonparametric generalizations of the general class of GLMM (Bush and MacEachern, 1996; Mukhopadhyay and Gelfand, 1997; Müller and Rosner, 1997; Walker and Mallick, 1997; Kleinman and Ibrahim, 1998a,b; Hanson, 2006; Jara et al., 2009). Bayesian nonparametric methods in item response theory (IRT) models have been also considered in the discussion of Ishwaran to Roberts and Rosenthal (1998), and in Duncan and MacEachern (2008), Karabatsos and Walker (2009) and Miyazaki and Hoshino (2009).

However, these developments and the associated applications have statistical sense only when the parameters of interest $(\beta, G)$ are identified by the observations. Recently, San Martín et al. (2010) have considered the identification problem of a nonparametric generalization of the Rasch model in a Bayesian setting. San Martín et al. (2010) showed that a semiparametric formulation of the Rasch model has not an empirical sense since the distribution $G$ generating the random effects is not identified by the observations, except in the unrealistic case of having a test of infinite length. Furthermore, when a finite number of items is available, they shown that only a finite number of properties of the general distribution $G$ can be identified. Consequently, the lack of identifiability of a general abilities distribution jeopardizes the empirical meaning of an estimate for $G$ under a finite number of items, which is of practical relevance, specially considering the large amount of research trying to relax the parametric assumption of $G$ in the
IRT literature or in more general GLMM. It should be mentioned that these conclusions are still valid for link functions \( F \) in (1.1) such that their support is finite. The only way out to the inferential problem in the Rasch model is, therefore, the comparisons among identified alternative parametric models.

The central topic of this paper is to provide an explicit proof of the identification of the parameters of interest in the context of Rasch-type models. These models are specified in a way similar to the Rasch model, but the logistic distribution in (1.3) is replaced by a general cumulative distribution function (cdf) \( F \), whereas the marginal model (1.2) corresponds to a family of distributions known up to the scale parameter. Thus, the central question to be discussed in this paper can be phrased as follows: under which conditions on the cdf’s \( F \) and \( G \), the item parameters \( \beta \) and the scale parameter \( \sigma \) are identified by the observations? It should be mentioned that our identification results complement or simplify some identification rules suggested in the psychometric and statistical literature.

A second topic considered in this paper is the extension of the identification analysis of Rasch-type models to two particular cases, namely when the item parameters are restricted or when the parameters of the distribution generating the random effects are restricted. Following De Boeck and Wilson (2004)’s terminology, these models are termed explanatory IRT-models at the item side and explanatory IRT-models at the person side, respectively. This topic is motivated by the fact that, in the psychometric literature, the identifiability of the parameters of interest in the statistical model (1.4) is typically obtained from the identification of the conditional model (1.3) (or (1.1)). However, as shown by San Martín et al. (2010), there not exist a relationship between the identification of the conditional model and the identification of the statistical model. In the context of explanatory IRT-models, we therefore show in which consist the difference between parameter identifiability in the statistical model and in the conditional model.

This paper is organized as follows: in Section 2, the identification strategy used in this paper is explained and illustrated. The identification analysis of Rasch-type models is developed in Section 3. In Section 4, these results are extended to some explanatory IRT-models, which are of interest in psychometrics. The paper concludes with a brief discussion.

## 2 Identification Strategy

### 2.1 The statistical model induced by descriptive IRT models

Let \( Y_p = (Y_{p1}, \ldots, Y_{pi})' \) denote the vector of all measurements available for person \( p \), where \( Y_{pi} = 1 \) if person \( p \) “correctly” answers item/probe \( i \), and \( Y_{pi} = 0 \) otherwise. Descriptive IRT models assume that the distribution of \( Y_{pi} \) conditionally on \((\theta_p, \beta_i)\) is given by

\[
P[Y_{pi} = 1 \mid \theta_p, \beta_i] = F(\theta_p - \beta_i), \quad p = 1, \ldots, N; \quad i = 1, \ldots, I,
\]

where \( F \) is a known continuous strictly increasing cumulative distribution function (cdf). If, for instance, \( F = \Phi \), then (2.1) corresponds to the probit model; if \( F \) is the logistic distribution, (2.1) corresponds to the Rasch model; see Rasch (1966b). It is also assumed that \((\text{H}1)\) for each person \( p \), his/her measurements are mutually independent conditionally on \((\theta_p, \beta)\), where \( \beta = (\beta_1, \ldots, \beta_I)' \) (this is the so-called
Axiom of Local Independence); (H2) the person-specific random effects $\theta_p$ are mutually independent and identically distributed, with a common distribution $G^\sigma$ defined as

$$P[\theta_p \leq x \mid \sigma] = G^\sigma((-\infty, x]) = G\left((-\infty, \frac{x}{\sigma})\right), \quad (2.2)$$

where $G$ is an arbitrary cumulative distribution function defined on $\mathbb{R}$, known up to the scale parameter $\sigma > 0$. The model is completed by assuming that (H3) $Y_{1}, \ldots, Y_{N}$ are mutually independent conditionally on $(\theta_1, \ldots, \theta_N, \beta)$.

The statistical model (or likelihood function), bearing on the observable variables only, is obtained after integrating out the person-specific random effects. It can be shown that $H1$, $H2$ and $H3$ jointly imply that the measurements $Y_{1}, \ldots, Y_{N}$ are mutually independent given $(\beta, \sigma)$, with a common probability distribution

$$P[Y_p = y_p \mid \beta, \sigma] = \int_{\mathbb{R}} \prod_{i=1}^{I} F(\theta - \beta_i)^{y_{pi}} [1 - F(\theta - \beta_i)]^{1-y_{pi}} G^\sigma(d\theta), \quad (2.3)$$

where $y_p = (y_{p1}, \ldots, y_{pI})' \in \{0, 1\}^I$. This statistical model is based on a hierarchical specification in which each person has his/her own unique effect (hypothesis H2), unexplained by person properties; and each item has its own unique effects (namely $\beta_i$), unexplained by item properties. Following De Boeck and Wilson (2004)'s terminology, the induced statistical model (2.3) is accordingly called structural-descriptive IRT-model (SD-IRT). A relevant characteristic of the statistical model (2.3) is that, for each person $p$, the measurements $Y_{p1}, Y_{p2}, \ldots, Y_{pI}$ are correlated provided that the link function $F$ is a strictly increasing monotonic function (see Sijtsma and Molenaar, 2002). This property is valid for each (non-degenerate) cdf $G$ generating the random effects.

### 2.2 Identification strategy

The identification of $(\beta, \sigma)$ by one observation $Y_p$ is entirely equivalent to their identification by an infinite sequence of measurements $Y_1, Y_2, \ldots$. This is due to the fact that, in the statistical model (2.3), the observations are mutually independent and identically distributed. The identification problem we are dealing with corresponds, therefore, to the injectivity of the mapping $(\beta, \sigma) \mapsto P[\cdot \mid \beta, \sigma]$, where $P[\cdot \mid \beta, \sigma]$ is given by (2.3). To solve this problem, we follow an identification strategy based on two steps:

1. A distinction between parameters of interest and identified parameters should be made. Such a distinction is based on the fact that a statistical model always involves an identified parametrization; for details, see San Martín et al. (2009, 2010).

2. An injective relationship (under restrictions, if necessary) between the parameters of interest and the identified parameters should be established. By so doing, the parameters of interest become identified by the observations.
In the case of the statistical model (2.3), the parameters of interest are \((\beta, \sigma)\). Now, the probabilities of the \(2^I\) different possible patterns are given by

\[
q_{12\ldots I} = P[Y_{p1} = 1, \ldots, Y_{p,I-1} = 1, Y_{pI} = 1 \mid \beta, \sigma]
\]

\[
q_{12\ldots I} = P[Y_{p1} = 1, \ldots, Y_{p,I-1} = 1, Y_{pI} = 0 \mid \beta, \sigma]
\]

\[
q_{12\ldots I} = P[Y_{p1} = 0, \ldots, Y_{p,I-1} = 0, Y_{pI} = 0 \mid \beta, \sigma].
\]

The statistical model (2.3) corresponds to a multinomial distribution \((Y_p \mid \pi) \sim \text{Mult}(2^I, \pi)\), where \(\pi = (q_{12\ldots I}, q_{12\ldots I-1}, \ldots, q_{1,2\ldots I})\). It is known that the parameter \(\pi\) of a multinomial distribution is identified. It corresponds, therefore, to the identified parameter indexing the statistical model (2.3); the \(q\)'s with less than \(I\) subscripts are linear combinations of them and, consequently, are also identified. Thus, the identifiability of the parameters of interest \((\beta, \sigma)\) follows after establishing an injectivity relation between them and functions of \(\pi\). This strategy is followed in the rest of the paper. Let us mention that this identification strategy has already been used in psychometrics (Muthen, 1979) and in biometrics (Rabe-Hesketh and Skrondal, 2001).

### 2.3 Illustration

Let us illustrate the previous identification strategy with the well-known probit model, which assumes that \(F = G = \Phi\), where \(\Phi\) is the cumulative standard normal distribution. In this case, the following \(I + 1\) equations can be established (for \(i = 1, \ldots, I\)):

\[
(i) \quad \alpha_i = P[Y_{pi} = 1 \mid \beta, \sigma] = \int_{\mathbb{R}} \Phi(\sigma \tau - \beta_i) \Phi(d\tau) = \Phi(-\delta_i);
\]

\[
(ii) \quad \alpha_{12} = P[Y_{p1} = 1, Y_{p2} = 1 \mid \beta, \sigma] = P[U_1 \leq -\delta_1, U_2 \leq -\delta_2],
\]

where

\[
\delta_i = \frac{\beta_i}{\sqrt{1 + \sigma^2}}, \quad i = 1, \ldots, I,
\]

and \((U_1, U_2)'\) is a normal random vector of expectation 0, the variances of \(U_1\) and \(U_2\) are equal to 1 and the covariance is given by

\[
cov(U_1, U_2) = \rho = \frac{\sigma^2}{1 + \sigma^2}.
\]

The identified parameters are \((\alpha_1, \ldots, \alpha_I, \alpha_{12})\) because they are functions of the identified parameter \(\pi\). Now, \(\delta_i\) for \(i = 1, \ldots, I\) are identified in view of the first equation. The second equation may be written as a function of \(\rho\), namely

\[
H(\rho) = P[U_1 \leq -\delta_1, U_2 \leq -\delta_2] = \int_{\mathbb{R}} \Phi(\frac{-\delta_2 - \rho U_1}{\sqrt{1 - \rho^2}}) \Phi(-\delta_1) dU_1.
\]
It may be shown that the derivative of $H(\rho)$ is strictly positive. Hence, $H(\rho)$ is strictly increasing on $(0,1)$ in $\rho$ and, therefore, $\sigma$ is identified. The identifiability of $\beta_i$ follows from (2.4).

3 Identification of Structural-Descriptive IRT models

3.1 Strategy of the identification analysis

In this section, we obtain the identification of $(\beta, \sigma)$ in the SD-IRT model (2.3) for arbitraries cdf’s $F$ and $G$. The identification analysis is based on the following two steps:

**STEP 1:** It is shown that the item parameters $\beta$ are a function of both the scale parameter $\sigma$ and the identified parameters $P[Y_{pi} = 1 | \beta, \sigma]$ for $i = 1, \ldots, I$.

**STEP 2:** It is shown that the scale parameter $\sigma$ is a function of the identified parameters $P[Y_{p1} = 1 | \beta, \sigma]$, $P[Y_{p2} = 1 | \beta, \sigma]$ and $P[Y_{p1} = 1, Y_{p2} = 1 | \beta, \sigma]$.

Let us remark that **STEP 1** corresponds to relation (2.4) in the illustration discussed in Section 2.3, whereas **STEP 2** corresponds to relation (2.5). Combining both steps, $(\beta, \sigma)$ becomes identified by one observation.

3.2 Proof of **STEP 1**

Assume that $F$ is a strictly increasing cdf. For all $i = 1, \ldots, I$, define

$$
\alpha_i \triangleq P[Y_{pi} = 1 | \beta, \sigma] = \int_{\mathbb{R}} F(\sigma x - \beta_i) G(dx),
$$

which is an identified parameter because it is a function of the identified parameter $\pi$ defined in Section 2.1. Clearly, the function

$$
p(\sigma, \beta) \triangleq \int_{\mathbb{R}} F(\sigma x - \beta) G(dx)
$$

is a continuous function in $(\sigma, \beta) \in \mathbb{R}_0^+ \times \mathbb{R}$ that is strictly decreasing in $\beta$ since $F$ is a strictly increasing continuous function; here $\mathbb{R}_0^+$ denotes the positive real line. Therefore, if we define

$$
\overline{p}(\sigma, \alpha) \triangleq \inf\{\beta : p(\sigma, \beta) < \alpha\},
$$

it follows that

$$
\overline{p}[\sigma, p(\sigma, \beta)] = \beta.
$$

Using (3.1), this implies that, for each $i = 1, \ldots, I$, $\beta_i = \overline{p}(\sigma, \alpha_i)$; that is, for each $i = 1, \ldots, I$, the item parameter $\beta_i$ is a function of both the scale parameter $\sigma$ and the identified parameter $\alpha_i$. 

6
3.3 Proof of Step 2

Suppose that \( I \geq 2 \) and let

\[
\alpha_{12} = P \{ Y_{p1} = 1, Y_{p2} = 1 \mid \beta, \sigma \} \tag{3.5}
\]

\[
= \int_{\mathbb{R}} F(\sigma x - \beta_1) F(\sigma x - \beta_2) G(dx).
\]

Using Step 1, the identified parameter \( \alpha_{12} \) can be written as a function of \( \sigma, \alpha_1 \) and \( \alpha_2 \), namely

\[
\alpha_{12} = q(\sigma, \alpha_1, \alpha_2) \tag{3.6}
\]

\[
= \int_{\mathbb{R}} F[\sigma x - \overline{p}(\sigma, \alpha_1)] F(\sigma x - \overline{p}(\sigma, \alpha_2)) G(dx).
\]

If \( F \) is a strictly increasing cdf with a continuous density function strictly positive on \( \mathbb{R} \), then it can be proved that the function \( \alpha_{12} = q(\sigma, \alpha_1, \alpha_2) \) is a strictly increasing continuous function of \( \sigma \). It follows that \( \sigma = \overline{q}(\alpha_{12}, \alpha_1, \alpha_2) \), where

\[
\overline{q}(\sigma, \alpha_1, \alpha_2) = \inf \{ \sigma : q(\sigma, \alpha_1, \alpha_2) > \alpha \}.
\]

In other words, \( \sigma \) becomes a function of identified parameters and, therefore, it is identified by the observations. The details are provided in Appendix A.

3.4 Main result

Summarizing, we obtain the following Theorem:

**Theorem 3.1** Consider the class of SD-IRT models (2.3), where \( F \) is a continuous strictly increasing cdf, with a continuous density function strictly positive on \( \mathbb{R} \). If at least two items are available, then the item parameters \( \beta \) and the scale parameter \( \sigma \) of the distribution \( G \) generating the person-specific random effects are identified by one observation \( Y_p \).

This theorem deserves the following comments:

1. Theorem 3.1 is valid for all distribution functions \( G^\sigma \) known up to the scale parameter \( \sigma \), including the discrete distributions.

2. Theorem 3.1 establishes the identification of both the item parameters and the scale parameter when \( F \) is a continuous strictly increasing distribution function, with a continuous density function strictly positive on \( \mathbb{R} \). This is the case when, for instance, the link function \( F \) is either a standard normal distribution, a logistic distribution or a complementary log-log distribution.

3. In particular, for the probit model, the identification of the scale parameter of the normal \( N(0, \sigma^2) \) is well established, proving thus a conjecture due to Chen and Dey (1998, p.352). For the logit link
function, Theorem 3.1 offers a rigorous proof of the identifiability of both the items parameters and the scale parameter in a structural Rasch model; it is particularly shown that, when $G$ is a normal distribution, the identification restriction corresponds to fix at zero the mean of the distribution generating the random effects.

4. Theorem 3.1 does not apply when an explanatory IRT-model is defined with a complementary-log link function defined as

$$F(\eta) = \begin{cases} 1 - \exp(-\eta), & \eta > 0, \\ 0, & \eta \leq 0. \end{cases}$$

In contrast to the above-mentioned link functions, this one may not be differentiable at $\eta = 0$ (see Piegorsch, 1992; Fahrmeir and Tutz, 2001), and its density function is positive on $\mathbb{R}^+$ only.

5. It should be remarked that Theorem 3.1 is still valid if all the items parameters are equal between them, that is, if $\beta_i = \beta$ for all $i = 1, \ldots, I$ and $I \geq 2$. This is not the case when the identification analysis is in the context of a fixed-effects specification of the Rasch model, that is, when $\theta_p$ is viewed as an unknown parameter and the statistical model (or likelihood function) is characterized by both (2.1) (with $F$ the logistic distribution) and the mutual independence of $\{Y_{pi} : p = 1, \ldots, N; i = 1, \ldots, I\}$. In this case, the parameters of interest are $(\theta_1, \ldots, \theta_N, \beta)$, which are identified if a linear restriction of the type

$$a'\beta = 0 \quad \text{for a known } a \in \mathbb{R}^I \text{ such that } a' I_I \neq 0 \quad (3.7)$$

is imposed; here $I_I = (1, \ldots, 1)' \in \mathbb{R}^I$. These conditions imply that $\beta \in \langle a \rangle^\perp$ and $I_I \notin \langle a \rangle^\perp$, where $\langle a \rangle$ denotes the linear space generated by the vector $a$. In other words, $\beta$ belongs to a linear space which excludes equal difficulties of all the items. Thus, in a fixed-effects specification of the Rasch model, the identification restriction not only fixes the scale of the parameters of interest, but also imposes a specific structure on the test in the sense that it should contain at least two items with different difficulties.

### 3.5 Identification when $G$ is known up to both the location and the scale parameters

Descriptive IRT-models can be related with explanatory IRT-models if the distribution generating the random effects is specified as a location-scale distribution $G^{\mu, \sigma}$ defined as

$$P[\theta_p \leq x \mid \mu, \sigma] = G^{\mu, \sigma}((-\infty, x]) \equiv G\left((-\infty, \frac{x - \mu}{\sigma})\right).$$

It is assumed that $G$ is known up to both the location parameter $\mu \in \mathbb{R}$ and the scale parameter $\sigma \in \mathbb{R}^+_0$. In this case, the parameters of interest are $(\beta, \mu, \sigma)$ and, consequently, the identification problem consists in identifying these parameters by the observations. This problem can be solved as a corollary of Theorem 3.1. As a matter of fact, the statistical model is now characterized by

$$P[Y_p = y_p \mid \beta, \mu, \sigma] = \int_{\mathbb{R}} \prod_{i=1}^{I} F(\theta - \beta_i)^{y_{pi}} [1 - F(\theta - \beta_i)]^{1-y_{pi}} G^\sigma(\theta) d\theta, \quad (3.8)$$
where \( y_p = (y_{p1}, \ldots, y_{pI})' \in \{0, 1\}^I \) and \( \beta_i \equiv \beta_i - \mu \). In particular, it holds that

\[
\tilde{\alpha}_i = \int_{\mathbb{R}} F(\sigma x - \tilde{\beta}_i) G(\, dx), \quad i = 1, \ldots, I, \quad (3.9)
\]

\[
\tilde{\alpha}_{12} = \int_{\mathbb{R}} F[\sigma x - p(\sigma, \tilde{\alpha}_1)] F(\sigma x - p(\sigma, \tilde{\alpha}_2)] G(\, dx),
\]

where \( p(\sigma, \tilde{\alpha}) \) is given by (3.3). These equations are similar to equations (3.1) and (3.6), which were derived in the context of a SD-IRT model when \( G \) is known up to the scale parameter \( \sigma \). Theorem 3.1 ensures, therefore, that \((\beta_1 - \mu, \ldots, \beta_I - \mu, \sigma)\) are identified by one observation. These identified parameters are in a one-to-one relationship with the parameters of interest \((\beta, \mu, \sigma)\) if a linear restriction of the form \( a' \beta = 0 \) such that \( \mathbb{I}_I a' \neq 0 \), with \( a \in \mathbb{R}^I \) a known vector, is assumed. By so doing, the parameters of interest become identified by the observations. It should be remarked that such an identification restriction excludes the case of equal difficulties of all the items. Summarizing, we obtain the following Corollary:

**Corollary 3.1** Consider the class of SD-IRT models, where \( F \) is a continuous strictly increasing cdf, with a continuous density function strictly positive on \( \mathbb{R} \), and the person-specific random effects are distributed according to a location-scale distribution \( G^{\mu, \sigma} \). If at least two items are available, then \((\beta, \mu, \sigma)\) are identified by one observation \( Y_p \) provided that \( a' \beta = 0 \) for a known \( I \)-dimensional vector \( a \) such that \( a' \mathbb{I}_I \neq 0 \).

### 3.6 Parameter interpretation for different specifications of the Rasch model

In the psychometric literature, the Rasch model is specified either considering the person ability \( \theta_p \) as a fixed effect (Rasch, 1966b), or considering it as a random effect (De Boeck and Wilson, 2004). Taking into account the identification analysis previously developed, it is possible to understand the statistical meaning of the corresponding parameters of interest, that is, their meaning with respect to the sampling process. Thus, the following conclusions can be drawn:

1. In a fixed-effects specification of the Rasch model, once an identification restriction has been chosen, the parameters of interest can be statistically interpreted. As a matter of fact, if \( \beta_1 = 0 \), then \( \theta_p \) represents the logarithm of the “betting odds” of a correct answer to the standard item 1, that is,

\[
\theta_p = \ln \left[ \frac{P(Y_{p1} = 1 \mid \theta_p, \beta_1 = 0)}{P(Y_{p1} = 0 \mid \theta_p, \beta_1 = 0)} \right].
\]

Similarly, the item parameter \( \beta_i \) corresponds to the logarithm of the odds ratio between item 1 and item \( i \) for each person \( p \), that is,

\[
\beta_i = \ln \left[ \frac{P(Y_{p1} = 1 \mid \theta_p, \beta_1 = 0)}{P(Y_{p1} = 0 \mid \theta_p, \beta_1 = 0)} \frac{P(Y_{pi} = 1 \mid \theta_p, \beta_i)}{P(Y_{pi} = 0 \mid \theta_p, \beta_i)} \right].
\]

The statistical meaning of \( \beta_i \) depends, therefore, not only on the item \( i \), but also on the standard item 1. For a discussion, see Rasch (1966a,b) and San Martín et al. (2009).
2. However, the statistical meaning of the parameters of interest \((\beta, \sigma)\) in a structural Rasch model is different from the previous one. As a matter of fact, as derived in Section 3.3, the scale parameter \(\sigma\) is given by \(q(\alpha_1, \alpha_2, \alpha_2)\), where \(\alpha_i\) (with \(i = 1, 2\)) corresponds to the marginal probability that each person \(p\) answers item \(i\) correctly (see equation (3.1)), and \(\alpha_{12}\) corresponds to the marginal probability that each person \(p\) answers both item 1 and item 2 correctly (see equation (3.5)). Thus, in a structural Rasch model, \(\sigma\) represents the dependency between items 1 and 2 induced by the marginal probabilities \(\alpha_1\) and \(\alpha_2\), and the joint probability \(\alpha_{12}\). Similarly, as derived in Section 3.2, the item parameter \(\beta_i\) is given by \(\beta_i = p(\sigma, \alpha_i)\) and, therefore, represents a function of both the marginal probability \(\alpha_i\) and the dependency between items 1 and 2 as measured by \(\sigma\). In other words, in a structural Rasch model, an item parameter not only capture information on the item itself, but also on items 1 and 2 through their dependency, whereas in a Rasch model specified under a fixed-effects approach, an item parameter provides information on both the item itself and the standard item without any kind of dependency between them.

3. When the distribution \(G\) generating the person-specific effects is known up to the scale and the location parameters, the statistical interpretation of both the item parameters \(\beta_i\) and the location parameter \(\mu\) depend on the linear restriction mentioned in Corollary 3.1. Thus, for instance, if \(\beta_1 = 0\), then \(\mu = -\bar{p}(\sigma, \bar{\alpha}_1)\), where \(\bar{\alpha}_1\) is given by (3.9); this implies that \(\mu < 0\) and that \(\beta_i = \bar{p}(\sigma, \bar{\alpha}_i) - \bar{p}(\sigma, \bar{\alpha}_1)\).

4 Identification of Explanatory-IRT-Models

Structural-explanatory IRT-models (SE-IRT models) correspond to restricted SD-IRT models. Three types of SE-IRT models can be specified: (i) \(SE\)-IRT models at the item side; (ii) \(SE\)-IRT models at the individual side; and (iii) doubly structural-explanatory IRT-models. The \(SE\)-IRT models at the item side are obtained by restricting the item parameters \(\beta\) of a SD-IRT model, whereas the \(SE\)-IRT models at the individual side are obtained by restricting the location parameter \(\mu\) and/or the scale parameter \(\sigma\) of a SD-IRT model. Doubly structural-explanatory IRT-models are obtained by combining the two previous specifications.

In this section, we show how the identification of the parameters of interest of SE-IRT models can be obtained from the identification of SD-IRT models. Furthermore, the difference between identification results obtained under a random-effects specification and under a fixed-effects specification is discussed.

4.1 Identification of structural-explanatory IRT-models at the item side: LLTM-type models

Following Fischer (1983) and De Boeck and Wilson (2004), in the linear logistic test model (LLTM), item properties are used to explain differences between items in terms of the effect they have on the probability to answer correctly the items. It is actually assumed that the difference between the item difficulty parameters \(\beta_i\) and \(\beta_j\) can only be due to structural characteristics of the respective items, that is, to the cognitive structure of the tasks underlying the solution of the item. This leads to restrict the
item parameters $\beta_i$ as

$$\beta_i = Q_i' \gamma + c, \quad i = 1, \ldots, I,$$

(4.1)

where $Q_i$ is a $K$-dimensional design vector. The entries of the vector $Q_i$ can be 0 and 1 (denoting the hypothetical absence or presence of an operation or task in solving item $i$), integers (denoting the hypothetical minimum number of times an operation or task has to be used in solving item $i$), or in general real numbers. The constant $c$ is viewed as an intercept (De Boeck and Wilson, 2004, p.62) and corresponds to the differences

$$\beta_i - \beta_j = \sum_{1 \leq k \leq K} [(Q_i)_k - (Q_j)_k] \gamma_k = c,$$

where $(Q_i)_k$ is the $k$-coordinate of the vector $Q_i$. This equality means that the difference of each of two item parameters is explained as the sum of the difficulty parameters of those cognitive operations which have to be performed in solving item $i$ but not in solving item $j$, and vice versa; see Fischer (1983, p.5). The specification of the model is completed by assuming that the person abilities are distributed according to a distribution $G^\sigma$ defined by (2.2).

In this context, the induced SE-IRT model is parameterized by $(\gamma, c, \sigma)$; these are the parameters of interest. Theorem 3.1 ensures, however, that $(\beta, \sigma)$ are identified by the observations, where $\beta = (\beta_1, \ldots, \beta_I)'$. Consequently, the parameters of interest $(\gamma, c, \sigma)$ become identified if an injective relationship between them and the identified parameters $(\beta, \sigma)$ is established. To do it, let $Q$ be a $I \times K$ matrix, where its $i$-th row is the vector $Q'_i$. Equality (4.1) can equivalently be rewritten as

$$(Q \mid I_I) \begin{pmatrix} \gamma \\ c \end{pmatrix} = \beta,$$

(4.2)

where $I_I = (1, \ldots, 1)' \in \mathbb{R}^I$. Therefore, an injective mapping between $(\beta, \sigma)$ and $(\gamma, c, \sigma)$ can be defined if the $r[(Q \mid I_I)] = K + 1$. This condition is equivalent to the following two conditions: (i) $r(Q) = K$; and (ii) $I_I \notin \text{Im}(Q)$, that is, $I_I$ does not belong to the linear space generated by the columns of $Q$. Summarizing, we obtain the following Corollary:

**Corollary 4.1** Consider SE-IRT models as specified in Section 2.1, where the link function $F$ is a continuous strictly increasing cdf, with a continuous density function strictly positive on $\mathbb{R}$; the person abilities are distributed according to a scale distribution $G^\sigma$; and the item parameters $\beta$ are restricted according to (4.2). Then

1. A necessary condition to identify $(\gamma, c, \sigma)$ by one observation is that $K \leq I$.

2. If at least two items are available, sufficient conditions to identify $(\gamma, c, \sigma)$ by one observation are that $r(Q) = K$ and $I_I \notin \text{Im}(Q)$. 

11
**Example 1** The arguments underlying Corollary 4.1 can be illustrated through the following example: consider $K = 3$ tasks and $I = 4$ items related according to the design matrix

$$Q = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$ 

Theorem 3.1 ensures that $(\beta_1, \beta_2, \beta_3, \beta_4, \sigma)$ are identified by $Y_p$. Now, the tasks parameters $\gamma_1, \gamma_2$ and $\gamma_3$, as well as the intercept $c$ are the parameters of interest and, therefore, should be identified by the observations. It can be verified that $r(Q) = 3$ and that $\mathbb{I}_4 \notin \text{Im} \ Q$; therefore, the matrix $(Q \ | \ \mathbb{I}_4)$ is non-singular and the parameters of interest $(\gamma_1, \gamma_2, \gamma_3, c)$ can be written as a function of the identified parameters $(\beta_1, \beta_2, \beta_3, \beta_4)$:

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ c \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & -1 \\ -1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} -\beta_2 + \beta_4 \\ \beta_1 + \beta_2 - \beta_3 - \beta_4 \\ \beta_1 - \beta_4 \\ -\beta_1 + \beta_3 + \beta_4 \end{pmatrix}.$$ 

Thus, $(\gamma_1, \gamma_2, \gamma_3, c)$ are identified because they are functions of identified parameters.

□

In the hierarchical specification considered in Corollary 4.1, the mean of the person-specific random effect is 0. An alternative specification is to consider the location parameter of the distribution $G$ equal to $c$ and, therefore, to omit the contribution of the constant predictor in the explanatory part of the model (see, e.g., De Boeck and Wilson, 2004, p.62); that is, to assume that

(i) $(\theta_p \ | \ c, \sigma) \overset{\text{iid}}{\sim} G^{c,\sigma}$,  
(ii) $\beta = Q\gamma$, 

(4.3)

where $Q$ is a $I \times K$ design matrix.

This hierarchical specification as well as that considered in Corollary 4.1 are equivalent models in the sense that they induce the same SE-IRT model; for a definition of equivalent models, see Maris and Bechger (2004). However, the identification analysis is not the same. As a matter of fact, if at least two items are available, Corollary 3.1 ensures that $(\beta, c, \sigma)$ are identified by one observation provided that a linear restriction of the form $a^T\beta = 0$, with $a$ a known $I$-dimensional vector such that $a^T\mathbb{I}_I \neq 0$. By (4.3.ii), this identification restriction is rewritten as $a^TQ\gamma = 0$. It remains to identify $\gamma$; from (4.3.ii), it follows that $\gamma$ is identified if $r(Q) = K$. Summarizing, we obtain the following Corollary:

**Corollary 4.2** Consider SE-IRT models as specified in Section 2.1, where the link function $F$ is a continuous strictly increasing cdf, with a continuous density function strictly positive on $\mathbb{R}$; the person abilities are distributed according to a location-scale distribution $G^{c,\sigma}$; and the item parameters $\beta$ are restricted as in (4.3.ii). Then

1. A necessary condition to identify $(\gamma, c, \sigma)$ by one observation is that $K \leq I$. 

12
2. If at least two items are available, sufficient conditions to identify \( (\gamma, c, \sigma) \) by one observation are that \( r(Q) = K \) and \( a' Q \gamma = 0 \) for a known \( I \)-dimensional vector \( a \) such that \( a' \mathbb{1}_I \neq 0 \).

**Example 2** Let us illustrate this corollary considering the design matrix \( Q \) of Example 1. In this case, the identified parameters are \( \tilde{\beta} = (\beta_1 - c, \beta_2 - c, \beta_3 - c, \beta_4 - c)' \), whereas the parameters of interest are \( (\gamma_1, \gamma_2, \gamma_3, c) \). Combining the linear restriction under which the mapping \( \tilde{\beta} \mapsto (\beta, c) \) is injective, along with restriction (4.3.ii), we obtain an equation from which the parameters of interest can be identified, namely

\[
\begin{pmatrix}
\gamma \\
c
\end{pmatrix} = \tilde{Q}^+ A^{-1} \begin{pmatrix}
\tilde{\beta} \\
0
\end{pmatrix},
\]

where

\[
A = \begin{pmatrix}
I_4 & \mathbb{1}_4' \\
\alpha' & 0
\end{pmatrix}, \quad \tilde{Q} = \begin{pmatrix}
Q & 0_4 \\
0_3' & 1
\end{pmatrix},
\]

\(0_4 = (0, 0, 0, 0)'\) and \( \alpha \in \mathbb{R}^4 \) is a known vector such that \( \alpha' \mathbb{1}_4 \neq 0 \). The latter condition ensures that \( A \) is a full rank matrix. Furthermore, \( r(\tilde{Q}) = r(Q) + 1 \). Therefore,

\[
\begin{pmatrix}
\gamma \\
c
\end{pmatrix} = \tilde{Q}^+ A^{-1} \begin{pmatrix}
\tilde{\beta} \\
0
\end{pmatrix},
\]

where \( \tilde{Q}^+ \) is the inverse of Moore-Penrose (Marsaglia, 1964). Thus, if for instance \( \alpha = (1, 0, 0, 0)' \), then

\[
\begin{pmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_3 \\
c
\end{pmatrix} = \begin{pmatrix}
\frac{2}{3} & -1 & -\frac{1}{3} & 2 & -\frac{1}{3} \\
-1 & 1 & 0 & 0 & 1 \\
-\frac{1}{3} & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\
-1 & 0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
\tilde{\beta}_1 \\
\tilde{\beta}_2 \\
\tilde{\beta}_3 \\
\tilde{\beta}_4
\end{pmatrix} = \begin{pmatrix}
\frac{1}{3}(2\beta_1 - 3\beta_2 - \beta_3 + 2\beta_4) \\
-\beta_1 + \beta_2 \\
\frac{1}{3}(-\beta_1 + 2\beta_3 - \beta_4) \\
-\frac{1}{2}\beta_1
\end{pmatrix}.
\]

\[\square\]

Corollaries 4.1 and 4.2 deserve the following comments:

1. The choice between using Corollary 4.1 or Corollary 4.2 may be arbitrary. However, if the design matrix \( Q \) is defined in such a way that \( \mathbb{1}_I \in \text{Im} (Q) \), as for instance

\[
Q = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1 \\
1 & 1 & 0
\end{pmatrix},
\]

then Corollary 4.1 does not apply and the identification of \( (\gamma, c, \sigma) \) follows from Corollary 4.2.
2. In Corollary 4.2, if the vector $\mathbf{a}$ is chosen in such a way that $\mathbf{a} \in \text{Ker} \left( Q' \right)$ (i.e., $\mathbf{a}$ belongs to the null space of matrix $Q'$), then the identification restriction $\mathbf{a}'Q\mathbf{\gamma} = 0$ is automatically satisfied provided that $\mathbb{1}_I \notin \text{Ker} \left( Q' \right) = \text{Im} \left( Q \right)$; for a proof of this equality, see Halmos (1987, section 49). As an example, consider the design matrix $Q$ of Example 1. We known that $\mathbb{1}_I \notin \text{Im} \left( Q \right)$. Furthermore, the Ker $\left( Q' \right)$ is generated by the vector $(1, 0, -1, -1)'$. Thus, if the vector $\mathbf{a}$ is chosen equal to $(1, 0, -1, -1)'$, then

$$
\begin{pmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_3 \\
c
\end{pmatrix} =
\begin{pmatrix}
0 & -1 & 0 & 1 & \frac{1}{3} \\
1 & 1 & -1 & -1 & -1 \\
0 & 0 & -1 & -\frac{2}{3} \\
-1 & 0 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
-\beta_2 + \beta_4 \\
\beta_1 + \beta_2 - \beta_3 - \beta_4 \\
\beta_1 - \beta_4 \\
\frac{1}{2}(\beta_1 + \beta_3 + \beta_4)
\end{pmatrix}
$$

3. Let us consider a fixed-effects specification of LLTM-type models, namely, that the individual ability $\theta_p$ is considered as an unknown parameter. In this case, the parameters of interest ($\theta_1, \ldots, \theta_N, \gamma_1, \ldots, \gamma_K, c$) are identified under two identification restrictions: (i) the item parameters $\beta$ satisfy a linear restriction of the type (3.7); (ii) $r(\left[ Q \mid \mathbb{1}_I \right]) = K + 1$ (which is equivalent to $r(Q) = K$ and $\mathbb{1}_I \notin \text{Im} \left( Q \right)$); see Fischer (1983, 1995, 2004). However, in a random-effects specification, Corollary 4.1 shows that it is possible to identifying the parameters of interest without introducing a linear restriction on the item parameters $\beta$.

4. Under a fixed-effects specification, Bechger et al. (2002) consider the identification problem of a LLTM-type model when the item parameters $\beta$ are restricted as $\beta = Q\mathbf{\gamma}$. In this context, $\mathbf{\gamma}$ is identified by the observations if the rank of the matrix $(I_I - \mathbb{1}_I \mathbf{a}')Q$ is equal to $K$, where the $I$-dimensional known vector $\mathbf{a}$ satisfies the equality $\mathbf{a}'\mathbb{1}_I = 1$. Under a random-effects specification, an argument similar to that of Corollary 4.1 leads to prove that $\mathbf{\gamma}$ is identified provided that $r(Q) = K$ only.

### 4.2 Identification of structural-explanatory IRT-models at the item side: MIRID-type models

Butter et al. (1998) (see also Smits et al., 2003) described the model with internal restrictions on item difficulties (MIRID) as a componential model for binary data. In the MIRID, the parameters of some items are defined to be a linear combination of the parameters of other items. The model requires that two sets of items be defined: composite items and component items. A composite item is an item that measures a concept that can be decomposed into components. A component item is an item that measures one of these components. The item parameters of the composite items are decomposed into parts attributed to the component items (the item parameters of the component items). As a simple example, Smits et al. (2003) propose the following one: $10 \times (5 + 3)$ as a composite item has two component items: $5 + 3$ and $10 \times 8$. The first component item is of the addition type; the second is of the multiplication type. For relevant psychological applications of the MIRID, see Smits and De Boeck (2003).

In general, $J$ item families ($j = 1, \ldots, J$) are defined so that, within each family, there is one composite item, to be conceived of as a dependent variable, and $K$ components items, to be conceived as the
independent variable. For the composite items, subscript $k$ is set to zero. The total number of items is $J(K + 1)$. The crucial assumption of the MIRID is that the item parameters of a composite item is a linear function of the item parameters of the associated component items:

$$
\beta_{j0} = \sum_{k=1}^{K} \omega_k \beta_{jk} + \tau \quad \text{for } j = 1, \ldots, J,
$$

(4.4)

where $\beta_{j0}$ is the item parameter of the composite item from item family $j$, $\beta_{jk}$ is the item parameter of the component item of type $k$ from item family $j$, $\omega_k$ is the weight of the component item parameters of type $k$ in determining the composite item parameters, and $\tau$ is a normalization constant.

Following Smits and De Boeck (2003), in an IRT-type model, each item (composite and component) has its own item parameter, so that the conditional probability that person $p$ will give a correct answer to item $jk$ is given by

$$
P[Y_{pjk} = 1 | \theta_p, \beta_{jk}] = F(\theta_p - \beta_{jk}), \quad j = 1, \ldots, J, \quad k = 0, \ldots, K.
$$

(4.5)

Hypotheses $H_1$, $H_2$ and $H_3$ as introduced in Section 2.1 are assumed to be valid for this class of models, and the random effects $\theta_p$'s are iid $G^\sigma$.

The corresponding SE-IRT model is indexed by $(\bar{\beta}, \omega, \tau, \sigma)$, where $\bar{\beta}$ is a $J(K + 1)$-dimensional vector containing the item parameters (composite and component), and $\omega = (\omega_1, \ldots, \omega_K)$. From Theorem 3.1 it follows that $(\bar{\beta}, \sigma)$ are identified by one observation if at least two items are available. It remains to identify $(\omega, \tau)$. But (4.4) can equivalently be rewritten as

$$
\beta_0 = (B | \mathbb{I}_K) \begin{pmatrix} \omega \\ \tau \end{pmatrix}
$$

(4.6)

where $\beta_0 = (\beta_{10}, \ldots, \beta_{J0})'$ and $B$ is a $J \times K$ matrix defined as

$$
B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{J1} & \cdots & \beta_{JK} \end{pmatrix},
$$

that is, a matrix of component item parameters. Therefore, $\omega$ and $\tau$ are identified if $r(B | \mathbb{I}_K) = K + 1$; in this case, $\omega$ and $\tau$ can be written as a function of the item parameters which in turn are identified by one observation. The rank identification restriction is equivalent to the following two conditions: $r(B) = K$ and $\mathbb{I}_K \notin \text{Im}(B)$. This implies that $J \geq K + 1$. Now, Theorem 3.1 requires that at least two items are available. It follows, therefore, that $J \geq 2$ and accordingly $K \geq 1$.

Summarizing, we obtain the following corollary:

**Corollary 4.3** Consider SE-IRT models specified as in Section 2.1, where the conditional probability that a person $p$ correctly answer the item $jk$ is given by (4.5); the link function $F$ is a continuous strictly increasing cdf, with a continuous density function strictly positive on $\mathbb{R}$; the person-specific random effects are distributed according to a scale distribution $G^\sigma$; and the item parameters $\beta$ are restricted according to (4.4). Then
1. A necessary condition to identify \((\tilde{\beta}, \omega, \tau, \sigma)\) by one observation is that \(J \geq K + 1\), with \(J \geq 2\) and \(K \geq 1\).

2. If \(J \geq 2\) and \(K \geq 1\), sufficient conditions to identify \((\tilde{\beta}, \omega, \tau, \sigma)\) by one observation are that \(r(B) = K\) and \(\mathbb{1}_K \notin \text{Im}(B)\).

This corollary deserves the following comments:

1. Corollary 4.3 shows that to obtaining the identification of \((\tilde{\beta}, \omega, \tau, \sigma)\) by the observations, it is not sufficient to fix the location of \(G\) at zero (when \(G = \Phi\), this is equivalent to fix the mean of the distribution of the person abilities at zero), but two additional identification restrictions should also be considered, namely \(r(B) = K\) and \(\mathbb{1}_K \notin \text{Im}(B)\). Thus, Corollary 4.3 complements the identification restriction suggested by Smits and Moore (2004, p. 272).

2. For the minimal case \(J = 2\) and \(K = 1\), the sufficient identification restrictions reduce to ensure that \(\beta_{11} \neq \beta_{21}\). In this case, \(\omega_1\) and \(\tau\) can easily be written as a function of the item parameters:

\[
\omega_1 = \frac{\beta_{10} - \beta_{20}}{\beta_{11} - \beta_{21}}, \quad \tau = \frac{\beta_{11}\beta_{20} - \beta_{21}\beta_{10}}{\beta_{11} - \beta_{21}}.
\]

As expected, \(\omega\) is actually an slope: if \(\omega > 1\) (resp. \(\omega < 1\)), then the distance between the two composite items is larger (resp. smaller) than the distance between the two component items.

3. The \(k\)-th column of matrix \(B\) is a \(J\)-dimensional vector whose coordinates correspond to the \(k\)-th component item of each family of composite items. Thus, for instance, if \(r(B) = K - 1\), then the first column \(\beta_{\bullet 1}\) can be written as a linear combination of the remaining columns of \(B\), that is, \(\beta_{\bullet 1} = \sum_{k=2}^{K} l_k \beta_{\bullet k}\), where at least one \(l_k\)'s is different from zero. In such a case, for each family of composite items, it follows that

\[
\beta_{j0} = \sum_{k=1}^{K} \omega_k \beta_k + \tau = \sum_{k=2}^{K} (\omega_k + l_k) \beta_k + \tau.
\]

Therefore, the identification restriction \(r(B) = K\) ensures that each composite item is decomposed into \(K\) component items only. This is also the meaning of the restriction \(\mathbb{1}_K \notin \text{Im}(B)\). In other words, the information provided by each component item is exhaustive in the sense that it is not recovered from the other component items, neither from the intercept.

4. In a fixed-effects specification of the MIRID, the parameters of interest are \((\theta, \tilde{\beta}, \omega, \tau)\), where \(\theta = (\theta_1, \ldots, \theta_N)'\). In this case, the item parameters \(\{\beta_{jk} : j = 1, \ldots, J; k = 0, 1, \ldots, K\}\) are identified under a linear restriction of the type (3.7); see Butter et al. (1998, p.51). Taking into account the relationship (4.6) (which is also valid under a fixed-effects specification of MIRID), the identification of \((\omega_1, \ldots, \omega_K, \tau)\) follows if \(r(B) = K\) and \(\mathbb{1}_K \notin \text{Im}(B)\). Thus, the identification of the parameters of interest is different if it is considered under a fixed-effects specification of MIRID-type models, or under a random-effects specification. In the former case, there also exist a
problem dealing with the scale of $\tau$, which is due to the linear restriction of the type (3.7) imposed on the item parameters; for details, see Butter et al. (1998, p.52), Smits and De Boeck (2003, pp.168-169) and Smits and Moore (2004, p.272). In a random-effects specification of MIRID-type models, this problem does not exist because the identifiability of the parameters of interest does not require a linear restriction of the type (3.7).

5. Maris and Bechger (2004) consider the identifiability of MIRID-type models under a fixed-effects specification. They specified relationship (4.5) in a way similar to LLTM-type models, namely $\beta = Q(\omega) \eta$, where $\omega$ represents the regression weights and $\eta$ contains both the parameters of the component items and the intercepts. More specifically, Maris and Bechger (2004) (see also Bechger et al., 2001) analyze the identification problem when $Q(\omega)$ can, after permuting rows and/or columns, be written as

$$
\begin{pmatrix}
I_{JK} & 0_{JK,m-JK} \\
A(\sigma)_{J,JK} & T_{J,m-JK}
\end{pmatrix};
$$

(4.7)

here, we are considering $J$ families of composite items, $K$ families of component items, and $m - JK$ intercepts. In this case, the parameters of interest are identified if (i) $\mathbb{I}_{JK} \neq \mathbb{I} \cdot \text{Im}(Q(\omega))$; (ii) $Q(\omega)$ is a matrix of full rank; and (iii) $Q(\omega_1)\eta_1 = Q(\omega_2)\eta_2$ implies that $(\omega_1, \eta_1) = (\omega_2, \eta_2)$; see Maris and Bechger (2004, Theorem 3, and Lemmas 1 and 2). It should be remarked that, when $m - JK = 1$, there exist specifications of the form $\beta = Q(\omega)\eta$ that can be rewritten in the form of equation (4.6). For instance, with six items consisting of four component items and two composite items, Butter et al. (1998, p.51) and Maris and Bechger (2004, p.628) consider the following design matrix:

$$
Q(\omega) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
\omega_1 & \omega_2 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & \omega_1 & \omega_2 & 1
\end{pmatrix},
$$

(4.8)

where $\eta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \tau)$, $\omega = (\omega_1, \omega_2)'$, $\beta_{11}, \beta_{12}, \beta_{21}$ and $\beta_{22}$ are the parameters of the component items, $\omega_1$ and $\omega_2$ are the regression weights, and $\tau$ is the intercept of the regression. The relationship between component items and composite items as summarized in matrix $Q(\omega)$ can be rewritten as follows:

$$
\begin{pmatrix}
\beta_{10} \\
\beta_{20}
\end{pmatrix} = \begin{pmatrix}
\beta_{11} & \beta_{12} & 1 \\
\beta_{21} & \beta_{22} & 1
\end{pmatrix} \begin{pmatrix}
\omega_1 \\
\omega_2 \\
\tau
\end{pmatrix}.
$$

Cases as the previous one can be analyzed under a random-effects specification; by so doing, the identification of the parameters of interest follows provided that $\tau(B) = K$ and $\mathbb{I}_{K} \notin \text{Im}(B)$. Thus, parameter identifiability under a random-effects specification is less restrictive than under a fixed-effects specification. Let us finally mention that, in the example we are discussing, the
parameters of interest are non identified under both a random-effects specification and a fixed-effects specification of the model; for a proof of the latter case, see Maris and Bechger (2004, Section 6.3).

4.3 Identification of structural-explanatory IRT-models at the person side

Let us consider explanatory IRT-models, where the item parameters are unrestricted, whereas the random-effects are distributed according to a location-scale distribution $G_{\mu,\sigma}$, with the location parameter $\mu$ restricted as

$$\mu_p = Z_p' \theta,$$

where $Z_p$ is a $L$-dimensional design vector and $\theta$ is an unknown parameter. This type of models are also known as latent regression IRT-models De Boeck and Wilson (2004, Chapter 2) and are typically used in the context of large-scale surveys as the PISA test (Wu, 2005).

The process generating the person-specific random effects is specified as follows:

(i) $(\theta_p | \mu_p, \sigma) \sim G_{\mu_p,\sigma}$ for each $p = 1, \ldots, N$;

(ii) $\theta_1, \ldots, \theta_N$ are mutually independent given $(\mu_1, \ldots, \mu_N, \sigma)$.

These conditions, along with the conditional mutual independence of $(Y_1, \ldots, Y_N)$ given $(\theta, \beta)$ (hypothesis $H3$), where $\theta = (\theta_1, \ldots, \theta_N)$, imply that the statistical model is characterized by the following conditions:

S1. $(Y_1, \ldots, Y_N)$ are mutually independent given $(\mu_1, \ldots, \mu_N, \beta, \sigma)$.

S2. The conditional distribution of $Y_p$ given $(\mu_1, \ldots, \mu_N, \beta, \sigma)$ depends on $(\mu_p, \beta, \sigma)$ only;

for a proof, see Mouchart and San Martín (2003). The statistical model is indexed by $(\beta, \theta, \sigma)$, which are the parameters to be identified by the observations.

By Corollary 3.1, it follows that $(\beta, \mu_p, \sigma)$ is identified by $Y_p$ for each $p = 1, \ldots, N$, provided a linear restriction on the item parameters of the type (3.7) is imposed. Under conditions S1 and S2, Theorem 2 in Mouchart and San Martín (2003) ensures that the previous $N$ identification restrictions imply the identifiability of $(\beta, \mu_1, \ldots, \mu_N)$ by $(Y_1, \ldots, Y_N)$.

It remains to identify $\theta$. From (4.9), it follows that

$$Z \theta = \mu,$$

where $\mu = (\mu_1, \ldots, \mu_N)'$, and $Z$ is a $N \times L$ matrix whose $p$-th row is $Z_p'$. Therefore, if $r(Z) = L$, $(\beta, \mu_1, \ldots, \mu_N)$ and $(\beta, \theta, \sigma)$ are related through an injective mapping. Summarizing, we obtain the following Corollary:

Corollary 4.4 Consider SE-IRT models specified as in Section 2.1, where the item parameters are unrestricted, the individual abilities are specified according to (4.10), the link function $F$ is a continuous strictly increasing cdf, with a continuous strictly positive density function on $\mathbb{R}$. Then
1. A necessary condition to identify \((\beta, \vartheta, \sigma)\) by one observation is that \(L \leq N\).

2. If at least two items are available, sufficient conditions to identify \((\beta, \vartheta, \sigma)\) by one observation are that \(r(Z) = L\) and that the item parameters satisfy a linear restriction of the type \(a' \beta = 0\) for a known \(I\)-dimensional vector \(a\) such that \(a' 1 \neq 0\).

5 Discussion

The identifiability problem is basic to the problem of statistical inference. Unless the parameters in a statistical model are identifiable, there is no meaning of estimability or estimation of such parameters as several combinations of different values for the parameters may lead to the same statistical model. The Bayesian approach does not offer a solution to this problem because identified parameters only are updated by the observations, and not the unidentified parameters (Florens et al., 1990; San Martín et al., 2010). Although this kind of statements are widely known in psychometrics, there does not exist a careful study of parameter identification in Rasch-type models when the person-specific parameters are viewed as random effects. In fact, parameter identification in the statistical model (2.3) is typically obtained from parameter identification in the conditional model (2.1). However, as it was shown by San Martín et al. (2010), there are no general relationships between the identification of the conditional model and the identification of the statistical model.

This paper provides an explicit proof of the identification of both the item parameters and the parameters indexing the distribution generating the random effects. This identifiability result is established in the statistical model (2.3). Moreover, it is compared with the parameter identification in the conditional model (2.1); the difference between this two types identification results is apparent, particularly in the explanatory IRT-models as developed in Section 4. Taking into account these results, it can be said that the parameter identification of other IRT-models, as the 2PL, the 1PL-G (San Martin et al., 2006), the 3PL, and more recently the 4PL (Loken and Rulison, 2010), are still open problems in spite of their wide use in applications. This paper intends to be an initial contribution to this theoretical discussion.

A Appendix

To prove that the function \(\alpha_{12}\) given by (3.5) is a strictly increasing continuous function of \(\sigma\), we need to study the sign of its derivative with respect to \(\sigma\). This requires not only to use the Implicit Function Theorem (see Spivak, 1965), but also to assume regularity conditions allowing to perform derivatives under the integral sign. We accordingly assume that the cdf \(F\) has a continuous density function \(f\) strictly positive on \(\mathbb{R}\). Furthermore, to prove that \(\alpha_{12}\) is a strictly increasing continuous function of \(\sigma\), we need to obtain the derivatives under the integral sign of the function \(p(\sigma, \beta)\) as defined in (3.2) with respect to \(\sigma\) and to \(\beta\). Consequently, the standard regularity conditions are of the type

\[
\int_{\mathbb{R}} f(\sigma x - \beta) G(dx) < \infty, \quad \int_{\mathbb{R}} |x| f(\sigma x - \beta) G(dx) < \infty,
\]
under which the Dominated Convergence Theorem can be applied to ensure the validity of performing derivatives under the integral sign.

Thus, under such regularity conditions, the function $p(\sigma, \beta)$ is continuously differentiable under the integral on $\mathbb{R}_0^+ \times \mathbb{R}$ and, therefore,

\begin{align*}
(i) \quad D_2 p(\sigma, \beta) & \doteq \frac{\partial}{\partial \beta} p(\sigma, \beta) = - \int_{\mathbb{R}} f(\sigma x - \beta) G(dx) \\
(ii) \quad D_1 p(\sigma, \beta) & \doteq \frac{\partial}{\partial \sigma} p(\sigma, \beta) = \int_{\mathbb{R}} x f(\sigma x - \beta) G(dx).
\end{align*}

(A.1)

Thus, $\overline{p}(\sigma, \alpha)$ as defined by (3.3) is also continuously differentiable on $\mathbb{R}_0^+ \times (0, 1)$ and from (3.4), we obtain that

\begin{align*}
(i) \quad 1 & = \frac{\partial}{\partial \beta} \overline{p}[\sigma, p(\sigma, \beta)] \\
& = D_2 \overline{p}[\sigma, p(\sigma, \beta)] \times D_2 p(\sigma, \beta) \\
(ii) \quad 0 & = \frac{\partial}{\partial \sigma} \overline{p}[\sigma, p(\sigma, \beta)] \\
& = D_1 \overline{p}[\sigma, p(\sigma, \beta)] + D_2 \overline{p}[\sigma, p(\sigma, \beta)] \times D_1 p(\sigma, \beta),
\end{align*}

(A.2)

where

$$D_1 p(\sigma, \alpha) \doteq \frac{\partial}{\partial \sigma} p(\sigma, \alpha), \quad D_2 p(\sigma, \alpha) \doteq \frac{\partial}{\partial \alpha} p(\sigma, \alpha).$$

Combining (A.1) and (A.2), we obtain that

\begin{align*}
(i) \quad D_2 \overline{p}(\sigma, \alpha) & = \frac{1}{D_2 p[\sigma, \overline{p}(\sigma, \alpha)]} = - \frac{1}{\int_{\mathbb{R}} f[\sigma x - \overline{p}(\sigma, \alpha)] G(dx)} \\
(ii) \quad D_1 \overline{p}(\sigma, \alpha) & = - \frac{D_1 p[\sigma, \overline{p}(\sigma, \alpha)]}{D_2 p[\sigma, \overline{p}(\sigma, \alpha)]} = \frac{\int_{\mathbb{R}} x f[\sigma x - \overline{p}(\sigma, \alpha)] G(dx)}{\int_{\mathbb{R}} f[\sigma x - \overline{p}(\sigma, \alpha)] G(dx)}
\end{align*}

(A.3)

where

$$P_{\sigma, \alpha}[X \in dx] \doteq G_{\sigma, \alpha}(dx) \doteq \frac{f[\sigma x - \overline{p}(\sigma, \alpha)] G(dx)}{\int_{\mathbb{R}} f[\sigma x - \overline{p}(\sigma, \alpha)] G(dx)}. \quad \text{(A.4)}$$

Thanks to the regularity conditions allowing to perform derivatives of $p(\sigma, \beta)$, and to the fact that $F \leq 1$, it can be shown that $\alpha_{12}$ is continuously differentiable under the integral sign in $\sigma$, $\beta_1$ and
\[ \beta_2; \text{ therefore the function } q(\sigma, \alpha_1, \alpha_2) \text{ is continuously differentiable under the integral sign with respect to } \sigma. \text{ It remains to show that the derivative w.r.t. } \sigma \text{ is strictly positive. Now, using (A.3.ii), we obtain that} \]

\[ \frac{\partial}{\partial \sigma} F[\sigma x - \bar{p}(\sigma, \alpha)] = (x - E_{\sigma, \alpha}(X)) f[\sigma x - \bar{p}(\sigma, \alpha)]. \quad \text{(A.5)} \]

But

\[ \int_{\mathbb{R}} (x - E_{\sigma, \alpha_1}(X)) f[\sigma x - \bar{p}(\sigma, \alpha_1)] F[\sigma x - \bar{p}(\sigma, \alpha_2)] G(dx) = \]

\[ = \int_{\mathbb{R}} f[\sigma x - \bar{p}(\sigma, \alpha_1)] G(dx) \times C_{\sigma, \alpha_1} \{ X, F[\sigma X - \bar{p}(\sigma, \alpha_2)] \}. \]

Now, since \( F[\sigma x - \bar{p}(\sigma, \alpha_2)] \) is a strictly increasing function of \( x \), the covariance between \( X \) and \( F[\sigma X - \bar{p}(\sigma, \alpha_2)] \) (with respect to \( G_{\sigma, \alpha_1} \) is strictly positive (if \( X \) is not degenerate). Furthermore,

\[ \int_{\mathbb{R}} f[\sigma x - \bar{p}(\sigma, \alpha_1)] G(dx) \]

is clearly strictly positive since the density function \( f \) is strictly positive on \( \mathbb{R} \). The two terms of the derivative of \( q(\sigma, \alpha_1, \alpha_2) \) are, therefore, strictly positive.

Acknowledgments: The first author was partially supported by the CEPPE CIEO-01 grant from the National Commission of Science and Technology of the Chilean Government, and by the PUENTE grant 08/2009 from the Pontificia Universidad Católica de Chile.

References


