INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0702

IDENTIFIABILITY AND ESTIMABILITY OF PARAMETRIC RASCH-TYPE MODELS

SAN MARTIN, E. and J.-M. ROLIN

Identifiability and Estimability of Parametric Rasch-Type Models

ERNESTO SAN MARTÍN[a,b] AND JEAN-MARIE ROLIN[c]

[a] *Department of Statistics, Pontificia Universidad Católica de Chile, Chile.*
[b] *Measurement Center MIDE UC, Pontificia Universidad Católica de Chile, Chile.*
[c] *Institut de statistique, Université catholique de Louvain, Belgium.*

February 13, 2007

Abstract

It is widely accepted that the Rasch model with abilities distributed according to a normal distribution is identified provided that the mean is fixed at zero. Nevertheless, the following questions arise: why is this result valid? Is it still valid if the distribution of the random effect (namely, the abilities) changes? Is it still valid if both the link function defining the item characteristic curve (ICC) and the distribution of the random effect change? What about the asymptotic behavior of the posterior expectation of identified parameters? Finally, what can we learn about model construction from these eventual identification results? This paper is intended to answer these questions.

Keywords: Sufficient parameter; Minimal sufficient parameter; Bayesian identifiability; Estimability; Monotonicity of the ICC; Generalized Linear Mixed Model.

1 Introduction

1.1 Rasch models specified as a fixed effect model

Rasch models are typically specified either as a fixed effect model or as a mixed effect model. The fixed model assumes that the sequence of binary random variables $\{Y_{ij} \ : \ 1 \leq i \leq n, 1 \leq j \leq m\}$ is mutually independent; here, $Y_{ij} = 1$ if subject i correctly answers item j. The response pattern of person i is written as $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})' \in \{0,1\}^m$. It is also assumed that the probability of the event $\{Y_{ij} = 1\}$ depends on two fixed effects, namely a subject effect $\theta_i \in \mathbb{R}$ and an item effect $\beta_j \in \mathbb{R}$, in such a way that

$$P[Y_{ij} = 1 \mid \theta_i, \beta_j] = F(\theta_i - \beta_j), \qquad i = 1, \ldots, n, \ \ j = 1, \ldots, m, \tag{1.1}$$

where F is the logistic distribution function. The function F is typically called *link function.* The subject effect θ_i represents the ability of subject i, whereas the item effect β_j corresponds to the difficulty of item j; see Rasch (1960). Using Neyman and Scott's (1948) terminology, the abilities correspond to *incidental parameters*, the dimension of which increases with the sample size, whereas the difficulties correspond to *structural parameters*.

As it is well known (see, in particular, Rasch, 1966), the parameter of interest $(\theta_1, \ldots, \theta_n, \beta_1, \ldots, \beta_m)$ is not identified by the observations. Since the pattern responses $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are mutually independent, the identifiability of the parameters of interest is obtained using *one* observation only. Thus, the lack of identifiability can be removed after introducing a linear restriction of the form $\boldsymbol{a}'\beta_1^m = 0$ such that $\boldsymbol{a}'\mathbb{1}_m \neq 0$, where $\boldsymbol{a} \in \mathbb{R}^m$, $\beta_1^m = (\beta_1, \ldots, \beta_m)'$ and $\mathbb{1}_m = (1, \ldots, 1)' \in \mathbb{R}^m$, under which the mapping $(\theta_1 - \beta_1, \ldots, \theta_1 - \beta_m) \longmapsto (\theta_1, \beta_1^m)$ becomes injective. The identification restriction excludes the case of *constant difficulties* among the items (*i.e.* $\beta_j = \beta$ for all $1 \leq j \leq m$). This identification analysis is entirely similar for all strictly increasing distribution function F.

1.2 Rasch models specified as a mixed effect model

Nevertheless, Rasch model is an example of the famous Neyman-Scott phenomenon (see Neyman and Scott, 1948; and Lancaster, 2000), namely that the maximum likelihood estimate (MLE) of the structural parameters is inconsistent due to the presence of the incidental parameters. Andersen (1973, 1980) proved the inconsistency of the MLE of β_j as $n \to \infty$ for $m = 2$ items, while Ghosh (1995) extended the proof for $m > 2$. Following Kiefer and Wolfowitz (1956), a way of recovering consistent estimates is to consider the abilities as a random effect in the sense that the θ_i's consist of independent random variables with a common probability distribution G_φ parameterized by φ. In this case, (1.1) should be viewed as a conditional model, whereas the process generating the abilities corresponds to the marginal latent model. More specifically, the Rasch model with random effects is defined by means of the following structural hypotheses:

H1. $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are mutually independent given $(\theta_1, \ldots, \theta_n, \beta_1^m)$.

H2. For each person i, the distribution of $\boldsymbol{Y}_i$ depends on (θ_i, β_1^m) only.

H3. For each person i, his/her responses $\{Y_{ij} : 1 \leq j \leq m\}$ are mutually independent given (θ_i, β_1^m). This corresponds to the Axiom of Local Independence.

H4. For each person i, Y_{ij} depends on (θ_i, β_j) for all $1 \leq j \leq m$; the specific functional dependency is given by (1.1).

H5. The abilities $\theta_1, \ldots, \theta_n$ are mutually independent given φ.

H6. $(\theta_i \mid \varphi) \sim G_\varphi$ for all $1 \leq i \leq n$, where $\varphi \in \mathbb{R}_+$ is a scale parameter, that is

$$P[\theta_i \leq x \mid \varphi] = G_\varphi([-\infty, x]) = G([-\infty, \frac{x}{\varphi}]). \tag{1.2}$$

The inference is based on the *structural Rasch model* (also called *statistical model*) obtained after integrating out the random effect θ_i. The structural hypotheses H1, H2 and H5 ensure that the patterns responses $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are mutually independent given (β_1^m, φ) with a common distribution given by

$$P[\boldsymbol{Y}_i = \boldsymbol{\epsilon}_i \mid \beta_1^m, \varphi] \;\; = \;\; \int_{\mathbb{R}} \prod_{j=1}^{m} F(\theta - \beta_j)^{\epsilon_{ij}} \left[1 - F(\theta - \beta_j)\right]^{1-\epsilon_{ij}} G_\varphi(d\theta), \tag{1.3}$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im}) \in \{0,1\}^m$; see, among others, Andersen and Madsen (1977), Bock and Aitkin (1981), Thissen (1982), De Leeuw and Verhelst (1986) and Molenaar (1995). Model (1.3) corresponds to a generalized linear mixed model (GLMM) with items as the main fixed effect and a random intercept, which defines the student's mastery level; see, among others, Mellenbergh (1994), Rijmen *et al.* (2003), De Boeck and Wilson (2004) and Tuerlinckx *et al.* (2006).

1.3 Identifiability of the structural Rasch model

Kiefer and Wolfowitz (1956) –see also Pfanzagl (1970)– proved that the MLE of the structural parameter –in this case, (β_1^m, φ)– is consistent provided it is identified. The question is to know under which conditions (β_1^m, φ) is identified. Different answers can be traced in the related literature. In the psychometric literature, sometimes it is assumed that G_φ is completely known (see, *e.g.*, Thissen, 1982, p. 176); the strict monotonicity of the distribution function F in (1.3) is, therefore, sufficient to identify the item parameters without additional restrictions. When G_φ is known up to a scale parameter, it is generally assumed (without proof) that the identification restriction needed to identify the Rasch model (1.1) (for instance, $\beta_1 = 0$) is sufficient to identify the structural parameters (β_1^m, φ) in model (1.3); see, among others, Adams, Wilson and Wang (1997), Smits and Moore (2004) and Rijmen and De Boeck (2005, p. 482). This approach means that the identifiability of the conditional model is sufficient for the identifiability of the statistical model. Nevertheless, such an implication is in general false, as the following counter-example shows: let $\boldsymbol{X}_i \in \mathbb{R}^2$ and $\omega = (a_0, \theta_1, \ldots, \theta_n, \Sigma)$, where Σ is a positive definite 2×2 symmetric matrix. Suppose that

$$(\theta_i \mid a_0, \Sigma) \overset{ind.}{\sim} \mathcal{N}(0,1), \qquad (\boldsymbol{X}_i \mid \omega) \overset{ind.}{\sim} \mathcal{N}_2\left(\begin{pmatrix} \theta_i \\ a_0\theta_i \end{pmatrix}, \Sigma \right) \tag{1.4}$$

and let (a_0, Σ) have a prior distribution equivalent to the Lebesgue measure. It is clear that ω is identified by $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ and, consequently, that (a_0, Σ) is identified by $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ conditionally on $(\theta_1, \ldots, \theta_n)$. If we marginalize this model with respect to θ_i, we obtain

$$(\boldsymbol{X}_i \mid a_0, \Sigma) \overset{ind.}{\sim} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} + 1 & \sigma_{12} + a_0 \\ \sigma_{12} + a_0 & \sigma_{22} + a_0^2 \end{pmatrix} \right), \tag{1.5}$$

loosing the identifiability of (a_0, Σ).

When the distribution of the random effect is a normal $\mathcal{N}(\mu, \varphi^2)$, it is stated that two different identification constraints can be imposed to obtain the identifiability of the structural parameter: either to fix the mean of θ at 0, or to set a particular β_i equal to 0; see Bechger *et al.* (2003, p. 328) and De Boeck and Wilson (2004, pp. 53-55) and the references mentioned there. On the other hand, Baker and Kim (2004) mention that "the identification problem is solved via the scaling of the posterior ability distribution" (p. 174). These approaches are widely accepted although their validity need to be formally established.

The statistical literature is not more precise with respect to the identification problem mentioned above. Chen and Dey (1998) pointed out that, when the distribution of the random effect is a normal $\mathcal{N}(0, \varphi^2)$, the variance φ^2 "is nearly not identified" (p. 325), but they do not provide a definition of "nearly identified". Swartz *et al.* (2004, p.3) state that any hierarchical model –an example of which is the Rasch model specified as a mixed effect model– is unidentified; for a similar comment, see Gelfand and Sahu (1999, p. 247).

1.4 Scope of the paper

The objective of this paper is to offer a clear bill about the identifiability of structural Rasch-type models. Section 3 offers a formal proof about the identifiability of the Normal Ogive (or, Probit) Rasch model (Lord, 1952) when the random effect is distributed according to a normal $\mathcal{N}(0, \varphi^2)$. The strategy is essentially based on a distinction between identified parametrization and structural parametrization, which is discussed in Section 2. Section 5 extends the identification analysis for a general link function F and a general distribution G_φ known up to a scale parameter φ, obtaining the identifiability of (β_1^m, φ) under general regularity conditions on F and G_φ. The case when G is known up to both a location parameter and a scale parameter is discussed as a corollary of the previous results. The identification analysis used in this section is mainly based on measurability arguments, which are described, in a general set-up, in Section 4. Section 6 briefly discusses the estimability (or Bayesian consistency) of the b-identified parameter in the class of structural Rasch-type models. The paper finishes with some concluding remarks.

2 Preliminaries

The structural Rasch model corresponds to a family of sampling distributions $P(\cdot \mid \beta_1^m, \varphi)$ defined by (1.3), which are indexed by the structural parameters $(\beta_1^m, \varphi) \in \mathbb{R}^m \times \mathbb{R}_+$. Let us remark that this parameter space is very different from the parameter space corresponding to the Rasch model specified as a fixed effect model, namely $\mathbb{R}^n \times \langle \mathbb{1}_m \rangle^\perp$, where $\langle \mathbb{1}_m \rangle$ is the subspace generated by $\mathbb{1}_m$. Consequently, the identification problem we are dealing with corresponds to analyze the injectivity of the mapping $(\beta_1^m, \varphi) \longmapsto P(\cdot \mid \beta_1^m, \varphi)$. In order to establish such an injectivity, it is useful to distinguish between a structural parametrization (which is not necessarily identified) and an identified parametrization. In the case of the structural Rasch model, the basic probabilities are the probability of the 2^m different possible patterns responses, namely

$$
\begin{aligned}
p_{12\cdots m} &= P[Y_{11}=1,\ldots,Y_{1,m-1}=1,Y_{1m}=1 \mid \beta_1^m,\varphi] \\
p_{12\cdots \bar{m}} &= P[Y_{11}=1,\ldots,Y_{1,m-1}=1,Y_{1m}=0 \mid \beta_1^m,\varphi] \\
&\vdots \\
p_{\bar{1}\bar{2}\cdots \bar{m}} &= P[Y_{11}=0,\ldots,Y_{1,m-1}=0,Y_{1m}=0 \mid \beta_1^m,\varphi].
\end{aligned}
$$

The p's with less than m subscripts are linear combinations of them. Therefore, the structural Rasch model can be completely described as a multinomial process, namely $(Y_i \mid \theta) \overset{\text{iid}}{\sim} \text{Mult}\,(2^m,\theta)$ for $1 \leq i \leq n$, where $\theta = (p_{12\cdots m}, p_{12\cdots m-1,\bar{m}},\ldots,p_{\bar{1},\bar{2},\ldots,\bar{m}})$. Since the mapping $\theta \longmapsto \text{Mult}(2^m,\theta)$ is injective (that is, the natural parameter of a multinomial distribution is identified), the identification of the structural parameter (β_1^m,φ) reduces to establish an injectivity between it and the identified parameter θ. In the next section, this strategy is used to study the identifiability of the structural parameter (β_1^m,φ) in the context of the structural normal ogive (or, probit) Rasch model.

3 Identifiability of the Structural Normal Ogive Rasch Model

Let us start by the Normal Ogive (or, Probit) Rasch Model introduced by Lord (1952). This model assumes that $F = G = \Phi$, where Φ is the cumulative standard normal distribution. In this case, we have that

$$
P\,[\,Y_{11} \,=\, 1 \mid \varphi,\beta_1\,] \,=\, \int_{\mathbb{R}} \Phi\,(\varphi\,\tau \,-\, \beta_1\,)\,\Phi\,(\,d\tau\,) \,=\, \Phi\left(\frac{-\,\beta_1}{\sqrt{1+\varphi^2}}\right) \tag{3.1}
$$

Since Φ is a strictly increasing function, it follows that $\dfrac{\beta_1}{\sqrt{1+\varphi^2}}$ is identified.

Now, considering the joint distribution of Y_{11} and Y_{12}, we obtain three equations with three unknown parameters β_1, β_2 and φ, namely

$$
\begin{array}{ll}
\text{(i)} & P\,[\,Y_{11} \,=\, 1 \mid \beta_1^m,\varphi\,] \,=\, P\,[\,U_1 \,\leq\, -\gamma_1\,] \,=\, \Phi\,(\,-\gamma_1\,) \\
\text{(ii)} & P\,[\,Y_{12} \,=\, 1 \mid \beta_1^m,\varphi\,] \,=\, P\,[\,U_2 \,\leq\, -\gamma_2\,] \,=\, \Phi\,(\,-\gamma_2\,) \\
\text{(iii)} & P\,[\,Y_{i1} \,=\, 1\,,\, Y_{i2} \,=\, 1 \mid \beta_1^m,\varphi\,] \,=\, P\,[\,U_1 \,\leq\, -\gamma_1\,,\, U_2 \,\leq -\gamma_2\,]
\end{array} \tag{3.2}
$$

where

$$
\gamma_i \,=\, \frac{\beta_i}{\sqrt{1\,+\,\varphi^2}} \qquad i \,=\, 1\,,\,2, \tag{3.3}
$$

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

$$C(U_1\,,\,U_2\,)\;=\;\rho\;\doteq\;\frac{\varphi^2}{1\,+\,\varphi^2}. \tag{3.4}$$

Clearly, γ_1 and γ_2 are identified in view of the two first equations. Now, the third equation may be written as a function of ρ, namely,

$$\begin{aligned} H\,(\,\rho\,) \;&\doteq\; P\,[\,U_1\,\leq\,-\gamma_1\,,\,U_2\,\leq\,-\gamma_2\,] \qquad (3.5)\\ &=\; E\,[\,P\,(\,U_2\,\leq\,-\gamma_2\,\mid\,U_1\,)\,\mathbb{1}_{\{\,U_1\,\leq\,-\gamma_1\,\}}\,]\;=\;E\left[\Phi\left(\frac{-\gamma_2\,-\,\rho\,U_1}{\sqrt{1\,-\,\rho^2}}\right)\mathbb{1}_{\{\,U_1\,\leq\,-\gamma_1\,\}}\right] \end{aligned}$$

It may be shown that the derivative of $H\,(\,\rho\,)$ is strictly positive. Hence, $H\,(\,\rho\,)$ is strictly increasing on $(\,0\,,\,1\,)$ in ρ, and therefore φ is identified. The same is true for β_1 and β_2 and the normal ogive (or, probit) model is then identified. We will give a general proof of this result in Section 5.

4 Identifiability and Measurability Arguments

The identification analysis of Section 3 is essentially based on the closed-form of equations (3.1) and (3.2). Unfortunately, this algebraic fact is in general not possible: take, for instance, the link function F as a logistic distribution and G as a standard normal distribution. Therefore, the identification strategy used in the previous section is not directly applied for general distribution functions F and G.

Nevertheless, the identification analysis of the normal ogive (or, probit) Rasch model provides insight to formalize an identification strategy based on measurability arguments. As a matter of fact, the main argument consists in establishing an injective relationship between the structural parameter (β_1^m, φ) and the identified parameter $\theta = (p_{12\cdots m}, p_{12\cdots m-1,\bar{m}}, \ldots, p_{\bar{1},\bar{2},\ldots,\bar{m}})$. As shown in Section 2, the identified parameter θ corresponds to the basic probabilities which describe the sample process. Therefore, *the identifiability of the structural parameters follows after writing them as a measurable function of sampling probabilities*. In what follows, we formalize these notions in terms of σ-fields; for readers unfamiliar with σ-field notions, we refer to Ellis and Junker (1997), particularly their appendix.

4.1 Sufficient parameter

Identifiability can be defined in terms of parametric minimal sufficiency; see, among others, Kadane (1974), Picci (1977) and Oulhaj and Mouchart (2003). We need, therefore, to review the concept of sufficiency and minimal sufficiency at the parameter level, already introduced in Barankin (1961) and in Barankin *et al.* (1980). A natural way to introduce the definition of sufficient parameter is by taking advantage of the symmetric role between parameters and observations in a Bayesian experiment (see, Florens *et al.*, 1990, chapter 1). Thus, the definition of sufficient parameter is equivalent to the definition of sufficient statistic by replacing “observations” by “parameters”. In a sampling context, a statistic

$S(X)$ is a sufficient statistics for θ if the conditional distribution of the sample X given $S(X)$ does not depend on θ. When θ is a random variable such a definition is equivalent to $X \perp\!\!\!\perp \theta \mid S(X)$.

Definition 4.1 *A function $g(\theta)$ of the parameter θ is a sufficient parameter for X if the conditional distribution of the sample X given θ is the same that the distribution of the sample X given $g(\theta)$, that is,*

$$X \perp\!\!\!\perp \theta \mid g(\theta). \tag{4.1}$$

Condition (4.1) implies that the distribution of X is completely determined by $g(\theta)$, or in other words, θ is redundant once $g(\theta)$ is known. By the symmetry of a conditional independence relation, it can also be concluded that $g(\theta)$ is a sufficient parameter if the conditional distribution of the redundant part θ given the sufficient parameter $g(\theta)$ is not updated by the sample, *i.e.*, $p(\theta \mid X, g(\theta)) = p(\theta \mid g(\theta))$.

Example 1 For the statistical model (1.5), the parameter (a_0, Σ) is a sufficient parameter.

Example 2 For the structural Rasch model (1.3), the parameter (β_1^m, φ) describe the sampling process generating the pattern responses $\boldsymbol{Y}_i$'s, that is, it is the corresponding sufficient parameter.

4.2 Minimal sufficiency and Bayesian identification

We might ask whether one sufficient parameter still contains redundant information or whether one sufficient parameter is any better than another. To see this point, consider a sufficient parameter $\phi \doteq g(\theta)$, *i.e.* $X \perp\!\!\!\perp \theta \mid \phi$. Note that the sample X does not increase the prior knowledge about θ given ϕ, or in other words, part of the prior information on θ is not revised by the sample. Therefore, the parametrization θ is not "identified" by the data X. This situation can be avoided if θ is a *minimal sufficient parameter*, that is, if θ is a sufficient parameter and it is a function of any other sufficient parameter. In this way, if the parametrization of a statistical model is based on a minimal sufficient parameter, then the parametrization does not contain redundant information. These considerations motivate the following definition (see Florens and Rolin, 1984):

Definition 4.2 *A sufficient parameter $\theta^* = h(\theta)$ is said to be Bayesian identified or b-identified by X, if θ^* is a minimal sufficient parameter for X.*

In other words, a parameter is said to be b-identified if it corresponds to the greatest possible parameter reduction for which the prior information is updated by the sample. Consequently, a b-identified parameter fully characterizes the learning process underlying a Bayesian model.

Since the posterior distribution of a non-identified parameter can always be computed, some Bayesian statisticians suggest that "unidentifiability cause no real difficulties in the Bayesian approach" (Lindley, 1971) and, therefore, that the inference can be based on such parameters. This perspective is followed, among others, by Leamer (1978), Poirier (1998), Gelfand and Sahu (1999), Ghosh *et al.* (2000) and Gustafson (2005). Nevertheless, as can be concluded from Definition 4.2, when a parameter is not identified, *part of the prior information is not revised by the observations*. Thus, if the parametrization of a statistical model is the minimal sufficient one, the parametrization does not contain redundant information.

Remark 1 *When both the parameter space and the sample space are Borel spaces, it is possible to demonstrate that if a parameter θ is identified in a sampling sense,* i.e.*, if the mapping $\theta \longmapsto P^{\theta}$ is injective, then it is b-identified for all prior distributions on the parameters; for details and proofs, see Florens* et al. *(1985) and Florens* et al. *(1990, chapter 4).*

4.3 Measurability with respect to sampling expectations

The minimal sufficient parameter of a Bayesian model is always almost surely a function of a countable number of sampling expectations; for a proof, see Appendix A. Therefore, a sufficient parameter θ is b-identified if there exists a measurable function h such that

$$\theta = h\left\{E[f(X) \mid \theta]\right\}, \tag{4.2}$$

for some measurable function f. However, to state this type of equalities is rather difficult, in particular in context of structural Rasch models. Fortunately, equality (4.2) can equivalently be stated in terms of measurability relationships: it is equivalent to say that θ is measurable with respect to the σ-field generated by the sampling expectations; see Neveu (1964, Proposition II.2.5). Denoting the relationship "g is measurable with respect to a σ-field $\mathcal{F}$" as "$g \in \mathcal{F}$", equality (4.2) is equivalent to

$$\theta \in \sigma\{E[f \mid \theta] \,:\, f \in \sigma(X)^{+}\}, \tag{4.3}$$

where $\sigma(Z)$ is the σ-field generated by a random variable Z and $\sigma(Z)^{+}$ denotes the set of positive measurable functions of Z. Taking into account that the σ-field generated by θ, $\sigma(\theta)$, is the smallest σ-field which makes θ a measurable function, then (4.3) is equivalent to

$$\sigma(\theta) \subset \sigma\{E[f \mid \theta] \,:\, f \in \sigma(X)^{+}\}. \tag{4.4}$$

Since $\sigma\{E[f \mid \theta] \,:\, f \in \sigma(X)^{+}\} \subset \sigma(\theta)$ by definition of conditional expectation, the relation (4.4) is equivalent to $\sigma(\theta) = \sigma\{E[f \mid \theta] \,:\, f \in \sigma(X)^{+}\}$. Heuristically, a σ-field $\sigma(Z)$ corresponds to the set of events that may be described in terms of that random variable (see, e.g., Florens *et al.*, 1993; and San Martín *et al.*, 2006). Therefore, relation (4.4) means that θ is b-identified if the information represented by it can be recovered from the information represented by the sampling expectations. Let us emphasize that the concept of b-identifiability is a genuinely Bayesian concept since it depends on the prior distribution through the prior null sets (*i.e.*, the events the prior probability of which is equal to 0 or to 1); see Appendix A.

Example 3 Continuing with the Example 1, the minimal sufficient parameter, which we denote by $\mathcal{A}^{*}$, is given by $\sigma\{E(f \mid a_0, \Sigma) \,:\, f \in \mathcal{B}^{+}_{\mathbb{R}^2}\}$, where $\mathcal{B}^{+}_{\mathbb{R}^2}$ denotes the set of positive functions measurable with respect to the Borel σ-field of $\mathbb{R}^2$. Let $\boldsymbol{X}_1 = (X_{11}, X_{12})'$; it follows that

$$\sigma_{11} + 1 = E(X_{11} \mid a_0, \Sigma) \in \mathcal{A}^*, \qquad \sigma_{22} + a_0^2 = E(X_{12} \mid a_0, \Sigma) \in \mathcal{A}^*,$$

$$\sigma_{12} + a_0 = E(X_{11}\, X_{12} \mid a_0, \Sigma) \in \mathcal{A}^*.$$

Therefore, $(\sigma_{11}, \sigma_{22} + a_0^2, \sigma_{12} + a_0)$ is the minimal sufficient parameter for $\boldsymbol{X}_1$; or, equivalently, it is the parameter b-identified by the observation $\boldsymbol{X}_1$. This characterization is valid for all the prior distributions on (a_0, Σ) equivalent to the Lebesgue measure.

5 Identifiability of the Structural Rasch Model for an arbitrary F and G

In the Rasch model defined by the structural hypotheses H1 to H6 (see Section 1.2), consider that the distribution G is known up to a scale parameter φ –see equation (1.2)–, and that F is a strictly increasing distribution function. The structural parameter (β_1^m, φ) is b-identified by $\boldsymbol{Y}_i$ if and only if

$$(\beta_1^m, \varphi) \in \sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in \sigma(\boldsymbol{Y}_i)^+\}, \tag{5.1}$$

where $\sigma(\boldsymbol{Y}_i)$ is the σ-field generated by $\boldsymbol{Y}_i$. Note that $\sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in \sigma(\boldsymbol{Y}_i)^+\}$ corresponds to the minimal sufficient parameter for $\boldsymbol{Y}_i$.

5.1 Main argument

Let $\mathcal{S}$ be the σ-field generated by the scale parameter φ, and $\mathcal{B}_j$ be the σ-field generated by β_j. In order to prove condition (5.1), let us suppose that the following two relations are true:

$$\beta_j \in \mathcal{S}\ \vee\ \sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in \sigma(Y_{ij})^+\} \subset \mathcal{S}\ \vee\ \sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in \sigma(\boldsymbol{Y}_i)^+\}; \tag{5.2}$$

and

$$\varphi \in \sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in (\,\sigma(Y_{i1}) \vee \sigma(Y_{i2})\,)^+\} \subset \sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in \sigma(\boldsymbol{Y}_i)^+\}, \tag{5.3}$$

where $\mathcal{F}_1 \vee \mathcal{F}_2$ denotes the smallest σ-field generated by $\mathcal{F}_1 \cup \mathcal{F}_2$. The subsets relationships in (5.2) and (5.3) are *always* true since $\sigma(Y_{ij}) \subset \sigma(\boldsymbol{Y}_i)$ and $\sigma(Y_{i1}) \vee \sigma(Y_{i2}) \subset \sigma(\boldsymbol{Y}_i)$, respectively. Relation (5.2) means that the difficulty parameter β_j is a measurable function of both the scale parameter φ and sampling expectations of the form $E[f(Y_{ij}) \mid \beta_j, \varphi]$, with f a Borel function defined on $\{0, 1\}$. Similarly, relation (5.3) tell us that the scale parameter φ is a measurable function of sampling expectations of the form $E[h(Y_{i1}, Y_{i2}) \mid \beta_1^m, \varphi]$, with h a Borel function defined on $\{0, 1\}^2$.

Now, relation (5.2) implies that

$$\sigma(\beta_1^m) \doteq \mathcal{B}_1^m \subset \mathcal{S}\ \vee\ \sigma\{E(f \mid \beta_1^m, \varphi)\ :\ f \in \sigma(\boldsymbol{Y}_i)^+\};$$

whereas relation (5.3) implies that

$$\mathcal{S} \subset \sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(\boldsymbol{Y}_i)^+\}.$$

Therefore,

$$\mathcal{B}_1^m \vee \mathcal{S} \;\subset\; \mathcal{S} \vee \sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(\boldsymbol{Y}_i)^+\} \;=\; \sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(\boldsymbol{Y}_i)^+\},$$

namely relation (5.1). In what follows, the basic relationships (5.2) and (5.3) are proved.

5.2 Proof of relation 5.2

For all $1 \le j \le m$ define

$$\alpha_j \doteq P\,[\,Y_{ij} = 1 \mid \beta_1^m, \varphi] = \int_{\mathbb{R}} F\,(\varphi\, x - \beta_j\,)\, G\,(\,dx\,), \tag{5.4}$$

which is measurable with respect to $\sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(Y_{ij})^+\}$. Clearly, the function

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

is a continuous function in $(\,\varphi\,,\,\beta\,) \in \mathbb{R}_0^+ \times \mathbb{R}$ that is strictly decreasing in $\beta \in \mathbb{R}$ since F is a strictly increasing continuous function. Therefore, if we define

$$\overline{p}\,(\,\varphi\,,\,\alpha\,) \doteq \inf\{\,\beta \,:\, p\,(\,\varphi\,,\,\beta\,) < \alpha\,\},$$

it is clear that

$$\overline{p}\,[\,\varphi\,,\,p\,(\,\varphi\,,\,\beta\,)\,] = \beta. \tag{5.5}$$

Using (5.4), this implies that, $\forall \quad 1 \le j \le m$,

$$\beta_j = \overline{p}\,(\,\sigma\,,\,\alpha_j\,) \in \mathcal{S} \vee \sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(Y_{ij})^+\},$$

namely relation (5.2).

Remark 2 *As mentioned in Section 4, the role of the identification analysis of the structural normal ogive Rasch model is to provide insights to generalize the proof for general structural Rasch-type models. It is, therefore, relevant to emphasize that, in the context of the structural normal ogive Rasch model, relation (5.2) corresponds to condition (3.3).*

5.3 Proof of relation 5.3

By making the following regularity assumptions on F and G, namely $F(x) \doteq \int_{-\infty}^{x} f(y)\,dy$, where f is a continuous strictly positive function on $\mathbb{R}$, and that, $\forall\, \varphi \in \mathbb{R}_0^+$ and $\forall\, \beta \in \mathbb{R}$, there exist $\epsilon > 0$ and $\eta > 0$, such that

$$\int_{\mathbb{R}} \max(|x|, 1) \sup_{|\varphi' - \varphi| \le \epsilon} \sup_{|\beta' - \beta| \le \eta} f(\varphi' x - \beta')\, G(dx) < \infty, \tag{5.6}$$

it is not difficult to prove that $p(\varphi, \beta)$ is continuously differentiable under the integral sign on $\mathbb{R}_0^+ \times \mathbb{R}$ and, therefore, that

$$\begin{aligned} &\text{(i)} \qquad D_2\, p(\varphi, \beta) \doteq \frac{\partial}{\partial \beta} p(\varphi, \beta) = -\int_{\mathbb{R}} f(\varphi x - \beta)\, G(dx) \\ &\text{(ii)} \qquad D_1\, p(\varphi, \beta) \doteq \frac{\partial}{\partial \varphi} p(\varphi, \beta) = \int_{\mathbb{R}} x\, f(\varphi x - \beta)\, G(dx). \end{aligned} \tag{5.7}$$

Therefore, $\overline{p}(\varphi, \alpha)$ is also continuously differentiable on $\mathbb{R}_0^+ \times (0, 1)$ and from (5.5), we obtain that

$$\begin{aligned} \text{(i)} \quad 1 &= \frac{\partial}{\partial \beta} \overline{p}[\varphi, p(\varphi, \beta)] \\ &= D_2\, \overline{p}[\varphi, p(\varphi, \beta)] \times D_2\, p(\varphi, \beta) \\ \text{(ii)} \quad 0 &= \frac{\partial}{\partial \varphi} \overline{p}[\varphi, p(\varphi, \beta)] \\ &= D_1\, \overline{p}[\varphi, p(\varphi, \beta)] + D_2\, \overline{p}[\varphi, p(\varphi, \beta)] \times D_1\, p(\varphi, \beta), \end{aligned} \tag{5.8}$$

where $D_1\, \overline{p}(\varphi, \alpha) \doteq \dfrac{\partial}{\partial \varphi} \overline{p}(\varphi, \alpha)$ and $D_2\, \overline{p}(\varphi, \alpha) \doteq \dfrac{\partial}{\partial \alpha} \overline{p}(\varphi, \alpha)$. Combining (5.7) and (5.8), we obtain that

$$\begin{aligned} \text{(i)} \quad D_2\, \overline{p}(\varphi, \alpha) &= \frac{1}{D_2\, p[\varphi, \overline{p}(\varphi, \alpha)]} = -\frac{1}{\int_{\mathbb{R}} f[\varphi x - \overline{p}(\varphi, \alpha)]\, G(dx)} \\ \text{(ii)} \quad D_1\, \overline{p}(\varphi, \alpha) &= -\frac{D_1\, p[\varphi, \overline{p}(\varphi, \alpha)]}{D_2\, p[\varphi, \overline{p}(\varphi, \alpha)]} = \frac{\int_{\mathbb{R}} x\, f[\varphi x - \overline{p}(\varphi, \alpha)]\, G(dx)}{\int_{\mathbb{R}} f[\varphi x - \overline{p}(\varphi, \alpha)]\, G(dx)} \\ &\doteq E_{\varphi, \alpha}(X), \end{aligned} \tag{5.9}$$

where

$$P_{\varphi\,,\alpha}\,[\,X\,\in\,dx\,]\;\doteq\;G_{\varphi\,,\alpha}\,(\,dx\,)\;\doteq\;\frac{f\,[\,\varphi\,x\,-\,\overline{p}\,(\,\varphi\,,\,\alpha\,)\,]\,G\,(\,dx\,)}{\int_{\mathbb{R}}\,f\,[\,\varphi\,x\,-\,\overline{p}\,(\,\varphi\,,\,\alpha\,)\,]\,G\,(\,dx\,)}. \tag{5.10}$$

Let us remark that the fact of writing $D_1\,\overline{p}(\varphi,\alpha)$ as an expectation with respect to the probability distribution (5.10) is relevant to prove the basic relation (5.3).

Now, let us consider

$$\begin{aligned}\alpha_{12} &\doteq P\,[\,Y_{i1}\,=\,1\,,\,Y_{i2}\,=\,1\,\mid\,\beta_1^m,\varphi\,]\\ &= \int_{\mathbb{R}} F\,(\varphi\,x\,-\,\beta_1\,)\,F\,(\varphi\,x\,-\,\beta_2\,)\,G\,(\,dx\,);\end{aligned}$$

this can be written as a function of φ, namely

$$\begin{aligned}\alpha_{12} &\doteq q\,(\,\varphi\,,\,\alpha_1\,,\,\alpha_2\,)\\ &= \int_{\mathbb{R}} F\,[\varphi\,x\,-\,\overline{p}\,(\,\varphi\,,\,\alpha_1\,)\,]\,F\,(\varphi\,x\,-\,\overline{p}\,(\,\varphi\,,\,\alpha_2\,)\,]\,G\,(\,dx\,).\end{aligned}$$

By definition, α_{12} is measurable with respect to $\sigma\{E(f\mid\beta_1^m,\varphi)\;:\;f\in(\,\sigma(Y_{i1})\vee\sigma(Y_{i2})\,)^+\}$. Furthermore, α_1 and α_2 are also measurable with respect to this σ-field. Therefore, if we show that the function $\alpha_{12}=q(\varphi,\alpha_1,\alpha_2)$ is a strictly increasing continuous function of φ, we obtain that

$$\varphi\,=\,\overline{q}\,(\alpha_{12}\,,\,\alpha_1\,,\,\alpha_2\,)\,\in\,\sigma\{E(f\mid\beta_1^m,\varphi)\;:\;f\in(\,\sigma(Y_{i1})\vee\sigma(Y_{i2})\,)^+\}, \tag{5.11}$$

namely relation (5.3), where

$$\overline{q}\,(\alpha\,,\,\alpha_1\,,\,\alpha_2\,)\,=\,\inf\{\,\varphi\,:\,q\,(\,\varphi\,,\,\alpha_1\,,\,\alpha_2\,)\,>\,\alpha\,\}.$$

Remark 3 *The monotonicity of α_{12} as a function of φ leads to obtain relation (5.3). In the context of the structural normal ogive Rasch model, this corresponds to the monotonicity of the function $H(\rho)$; see equation (3.6).*

Thanks to assumption (5.6) and the fact that $F\;\leq\;1$, it can be shown that α_{12} is continuously differentiable under the integral sign in φ, β_1 and β_2, and therefore the function $q\,(\,\varphi\,,\,\alpha_1\,,\,\alpha_2\,)$ is continuously differentiable under the integral sign with respect to φ. It remains to show that this derivative is strictly positive. Now, using (5.9.ii), we obtain that

$$\frac{\partial}{\partial\,\varphi}\,F\,[\varphi\,x\,-\,\overline{p}\,(\,\varphi\,,\,\alpha\,)\,]\,=\,(\,x\,-\,E_{\varphi\,,\alpha}\,(\,X\,)\,)f\,[\varphi\,x\,-\,\overline{p}\,(\,\varphi\,,\,\alpha\,)\,]. \tag{5.12}$$

But

$$\int_{\mathbb{R}} (x - E_{\varphi,\alpha_1}(X)) f[\varphi x - \overline{p}(\varphi, \alpha_1)] F[\varphi x - \overline{p}(\varphi, \alpha_2)] G(dx)$$
$$= \int_{\mathbb{R}} f[\varphi x - \overline{p}(\varphi, \alpha_1)] G(dx) \times C_{\varphi,\alpha_1}\{X, F[\varphi X - \overline{p}(\varphi, \alpha_2)]\}$$

Now, since $F[\varphi x - \overline{p}(\varphi, \alpha_2)]$ is a strictly increasing function of x, the covariance between X and $F[\varphi X - \overline{p}(\varphi, \alpha_2)]$ (with respect to G_{φ,α_1}) is strictly positive (if X is not degenerate). Furthermore, $\int_{\mathbb{R}} f[\varphi x - \overline{p}(\varphi, \alpha_1)] G(dx)$ is clearly strictly positive. The two terms of the derivative of $q(\varphi, \alpha_1, \alpha_2)$ are, therefore, strictly positive.

5.4 Main result

Summarizing, we obtain the following theorem:

Theorem 5.1 *Consider the family of Rasch-type models specified by the structural hypotheses H1 to H6 introduced in Section 1.2. The difficulty parameters β_1^m and the scale parameter φ of the distribution generating the abilities are b-identified by $\boldsymbol{Y}_1$ if the following two conditions hold:*

(i) The link function F defining the ICC is a strictly increasing continuous function with a continuous strictly positive density function.

(ii) The distribution functions F and G satisfy the regularity condition (5.6).

In particular, the structural Normal Ogive Rasch model and the structural Rasch model (*i.e.*, when F corresponds to the logistic distribution and $G = \Phi$) satisfy the hypotheses of this theorem.

Remark 4 *The identification analysis developed in this section can also be considered as a classical analysis in the sense that the the main arguments reduce to establish measurability relationships between the structural parameters and the identified ones. This requires to endow the parameter space with a σ-field in such a way that the sampling probabilities become transitions probabilities; for details, see Caillot and Martin (1972), Florens* et al. *(1985) and Florens* et al. *(1990, section 4.6.2). Consequently, Theorem 5.1 is also valid in a sampling set-up.*

The identification restrictions established in Theorem 5.1 does not exclude constant difficulty (*i.e.*, $\beta_j = \beta$ for all $j = 1, \ldots, m$) as it is the case for the identification restriction of model (1.1), the parameter space of which is, for instance, given by

$$(\theta_1^n, \beta_1^m) \in \mathbb{R}^n \times \langle \mathbb{1}_m \rangle^\perp; \tag{5.13}$$

see Section 1.1. Nevertheless, if the distribution G generating the random effects is known up to both a location parameter μ and a scale parameter φ, then it is necessary to impose an identification restriction on the difficulty parameters, which leads to exclude constant difficulties. As a matter of fact, let $G_{\mu,\varphi}$ be a probability distribution given by

$$P[\theta_i \leq x \mid \varphi, \mu] = G([-\infty, \frac{x-\mu}{\varphi}]).$$

Relation (5.4) is, therefore, rewritten as

$$\alpha_j \doteq P\,[\,Y_{ij} \,=\, 1 \,\mid\, \beta_1^m, \varphi, \mu\,] \,=\, \int_{\mathbb{R}} F\,(\varphi\, x \,+\, \mu \,-\, \beta_j\,)\, G\,(\,dx\,) \qquad \forall\, 1 \leq j \leq m.$$

Since F is a strictly increasing continuous function, we have, for all $1 \leq j \leq m$,

$$\beta_j - \mu \in \mathcal{S} \vee \sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(Y_{ij})^+\} \subset \mathcal{S} \vee \sigma\{E(f \mid \beta_1^m, \varphi) \,:\, f \in \sigma(\boldsymbol{Y}_i)^+\}.$$

Using the same arguments developed in Sections 5.1, 5.2 and 5.3, it can be concluded that $(\beta_1 - \mu, \ldots, \beta_m - \mu, \varphi)$ is b-identified by $\boldsymbol{Y}_1$. Therefore, under a restriction of the form $\boldsymbol{a}'\beta_1^m = 0$ such that $\boldsymbol{a}'\mathbb{1}_m \neq 0$, the structural parameter $(\beta_1^m, \mu, \varphi)$ is b-identified by $\boldsymbol{Y}_1$. In this case, the parameter space of the structural Rasch model with a distribution $G_{\mu,\varphi}$ known up to both a location parameter μ and a scale parameter φ is, for instance, given by

$$(\beta_1^m, \mu, \varphi) \;\in\; \langle \mathbb{1}_m \rangle^{\perp} \;\times\; \mathbb{R} \;\times\; \mathbb{R}_+, \tag{5.14}$$

which, although different from the parameter space (5.13), also excludes constant difficulties.

Theorem 5.1 also provides some insight at the model construction level. As a matter of fact, the strict monotonicity of the continuous distribution function F defining the ICC is a *necessary* identification condition. As well known, this condition is typically assumed in nonparametric IRT specifications (see, *e.g.*, Ellis and Junker, 1997; and Sijtsma and Molenaar, 2002), and when the Rasch model is constructively deduced (see, *e.g.*, Fischer, 1995). Nevertheless, when modeling guessing responses, non-monotonic ICCs have been suggested as a "desirable feature", the idea being that "examinees with sufficiently low θ can only guess randomly, examinees with higher θ may be misinformed and may do less well than a random guess" (Lord, 1983, pp. 479-480); see also Mislevy and Bock (1982); for a summary of different positions, see Hutchinson (1991, sections 2.7.2 and 2.10.7). These models fail to be identified either if we consider the conditional model given the abilities, or the statistical model obtained after integrating out the abilities.

6 Estimability of the Structural Rasch-Type Models

Let θ be a parameter to be estimated. Since the objective of Bayesian inference is the transformation of "prior to posterior" distributions, Bayesian estimators are typically based on the posterior distribution.

The posterior expectation of a parameter of interest $f(\theta)$ *always* converges almost surely to $E[f(\theta) \mid X_1^\infty]$, the posterior expectation of θ given the infinite sequence of observations, provided that $f(\theta)$ be an integrable function; this is due to the Martingale Theorem. By analogy to sampling consistency, it seems natural to say that the Bayesian estimator $E[f(\theta) \mid X_1^n]$ is consistent if its limit is a.s. equal to $f(\theta)$, that is, $E[f(\theta) \mid X_1^\infty] = f(\theta)$ a.s.; here $X_1^n = (X_1, \ldots, X_n)$. These facts motivate the following definition:

Definition 6.1 *Let f be an integrable function. The parameter $f(\theta)$ is said to be estimable if the posterior expectation $E[f(\theta) \mid X_1^n]$ converges a.s. to $f(\theta)$.*

Using the basic properties of the conditional expectation, the condition $E[f(\theta) \mid X_1^\infty] = f(\theta)$ a.s. is equivalent to say that $f(\theta)$ is a.s. a function of X_1^∞. This means that the infinite sequence of observations X_1^∞ contains the relevant information necessary to *construct* estimable parameters. This condition provides a Bayesian interpretation about the meaning of a parameter.

In an iid process, the estimability (or, Bayesian consistency) of identified parameters follow from the following general theorem:

Theorem 6.1 *Let $\{X_n : n \in \mathbb{N}\}$ be an iid process conditionally on θ. Then the corresponding b-identified parameter θ^* is consistently estimated by $E(\theta^* \mid X_1^n)$.*

For a proof, see Florens and Rolin (1984) and Florens *et al.* (1990, chapter 9).

As pointed out in Section 1.2, hypotheses H1, H2 and H5 imply that the sequence $\{\boldsymbol{Y}_i : i \in \mathbb{N}\}$ forms an iid process conditionally on (β_1^m, φ). Therefore, as a corollary of Theorem 6.1, we obtain the following proposition:

Proposition 6.1 *Consider the family of Rasch-type models specified by the structural hypotheses H1 to H6 introduced in Section 1.2, such that the identification restrictions established in Theorem 5.1 are satisfied. It follows that*

$$\lim_{n\to\infty} E[h(\beta_1^m, \varphi) \mid \boldsymbol{Y}_1^n] = h(\beta_1^m, \varphi) \quad \textit{a.s.}$$

for all integrable function h, where $\boldsymbol{Y}_1^n = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$.

This proposition establishes not only that both each difficulty parameter β_j's and the scale parameter φ of the distribution generating the abilities are estimable, but also that all the functions of any combination of them are consistently estimated. Furthermore, it can be proved that in an iid process the minimal sufficient statistics corresponding to the asymptotic experiment $(\boldsymbol{Y}_1^\infty \mid \beta_1^m, \varphi)$ is a.s. equal to a function of the b-identified parameter (β_1^m, φ); for details, see Florens *et al.* (1990, chapter 9). Therefore, in the asymptotic Bayesian experiment characterized by the random variables $(\beta_1^m, \varphi, \boldsymbol{Y}_1^\infty)$, Bayesian and sampling estimates provide the same information. This result is in agreement with Kiefer and Wolfowitz (1956) and Pfanzagl (1970).

7 Concluding Remarks

As it is well known, the identification problem was couched in the context of structural modeling, becoming a "necessary part of the specification problem" (Koopmans and Reiersøl, 1950, p. 169). The basic idea of structural modeling is to offer a substantive explanation of an observed phenomenon. Technically, the probability describing the observed variables is obtained from a hierarchical model after integrating out the marginal model generating the latent variables. The hierarchical structure is expected to be motivated by substantive considerations. Thus, the concern underlying identifiability is to know if the observed phenomenon is generated by *only one structure*; otherwise, "the first explanation of the data is not the only one" (Clifford, 1982); see also Hurwicz (1950), Clifford (1982a) and Manski (1995).

Rasch models, and by extension the GLMMs, constitute a relevant example of structural modeling. It is, therefore, surprising that the related literature does not pay a careful attention to the identification problem. Taking into account the discussion of Section 1.3, four statements need to be evaluated: (i) the identifiability of the structural Rasch model (1.3) follows from the identifiability of the conditional model (1.1); (ii) the identifiability of the structural Rasch model can be solved via the scaling of the posterior ability distribution; (iii) from a Bayesian point of view, hierarchical models are unidentified; (iv) the identifiability of the parameter scale φ of the distribution G generating the random effects is not well established. These statements can actually be verified once we pay attention to the binary character of the identification concept, namely which parameter we intend to identify by which observation; see Definition 4.2. By so doing, the following conclusions were established in this paper:

1. The identifiability of the Rasch model at the conditional level (1.1) is different from the identifiability of the structural Rasch model (1.3). This is due to the fact that the corresponding parameter spaces are different. Therefore, the identifiability of structural Rasch models should formally be stated, which is done in Theorem 5.1. This qualifies statement (i) above.

2. The identification analysis is mainly focused on the sampling distributions up to prior null sets. This qualifies statement (ii) above.

3. Rasch-type models are an example of hierarchical models or GLMMs. We have proved that these models are identified from a Bayesian point of view. This qualifies statement (iii) above.

4. The scale parameter φ of the distribution G is identified by one observation since it is a measurable function of sampling expectations; see the basic relation (5.3). This qualifies statement (iv) above.

The identification problem analyzed in this paper need to be studied in the context of 2PL and 3PL models with random effects. Moreover, taking into account that the distribution G generating the abilities is a parameter of interest, it is relevant to know if G has an empirical, that is, if G is identified in the context of a structural Rasch model. In a next paper these authors intend to address this problem.

Appendix

A A characterization of the minimal sufficient parameter

Let us consider a Bayesian experiment $\mathcal{E} = (A \times Y, \mathcal{A} \vee \mathcal{Y}, \Pi)$, where $(A, \mathcal{A})$ (resp., $(Y, \mathcal{Y})$) is the *parameter space* (resp., the *sample space*), and Π is a unique probability measure defined on $\mathcal{A} \vee \mathcal{Y}$. Let $\Sigma_{\mathcal{E}}$ be the class of sufficient parameters $\mathcal{B} \subset \mathcal{A}$ for the process generating $\mathcal{Y}$, namely $\Sigma_{\mathcal{E}} = \{\mathcal{B} \subset \mathcal{A} : \mathcal{Y} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}\}$. It follows that $\mathcal{A} \in \Sigma_{\mathcal{E}}$, hence $\Sigma_{\mathcal{E}} \neq \emptyset$. Therefore, if $\mathcal{B}_1, \mathcal{B}_2 \in \Sigma_{\mathcal{E}}$, then $\overline{\mathcal{B}_1} \cap \overline{\mathcal{B}_2} \in \Sigma_{\mathcal{E}}$, where $\overline{\mathcal{B}_j}$ $(j = 1, 2)$ denotes the completion $\overline{\mathcal{B}_j} = \mathcal{B}_j \vee \{E \in \mathcal{A} : \mu(E)^2 = \mu(E)\}$ and μ denotes the restriction of Π on $\mathcal{A}$ (that is, the prior distribution). Consequently, the minimal sufficient parameter $\mathcal{B}_{\text{min}} \in \Sigma_{\mathcal{E}}$ *always* exists and it is given by

$$\mathcal{B}_{\text{min}} = \bigcap_{\mathcal{B} \in \Sigma_{\mathcal{E}}} \overline{\mathcal{B}}.$$

Using the properties of the measurable completion (see Florens *et al.*, 1990, chapter 2), it can be verified that $\overline{\mathcal{B}_{\text{min}}} = \mathcal{B}_{\text{min}}$. Thus, the minimal sufficient parameter $\mathcal{B}_{\text{min}}$ contains *all* the null sets of the parameter space $(A, \mathcal{A})$ defined with respect to the prior probability μ.

The minimal sufficient parameter $\mathcal{B}_{\text{min}}$ can be expressed in more operational terms, namely as a σ-field generated by sampling expectations. Indeed, by definition of a σ-field generated by a function, the σ-field generated by every version of the sampling expectations, namely $\sigma\{E(f \mid \mathcal{A}) : f \in \mathcal{Y}^+\}$, is the smallest sub-$\sigma$-field of $\mathcal{A}$ that makes the sampling expectations measurable; here $\mathcal{Y}^+$ is the set of $\mathcal{Y}$-measurable positive functions. Then:

1. By construction of $\mathcal{B}_{\text{min}}$, it follows that $\mathcal{A} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{B}_{\text{min}}$. Since the conditional independence property corresponds to a measurability condition (see Theorem 2.2.6 in Florens *et al.*, 1990), it follows that $E\,(f \mid \mathcal{A}) \in \mathcal{B}^+_{\text{min}} \quad \forall\, f \in \mathcal{Y}^+$. Therefore, $\sigma\{E(f \mid \mathcal{A}) : f \in \mathcal{Y}^+\} \subset \mathcal{B}_{\text{min}}$.

2. Similarly, $E(f \mid \mathcal{A}) \in \sigma\{E(f \mid \mathcal{A}) : f \in \mathcal{Y}^+\}$, so $\mathcal{A} \perp\!\!\!\perp \mathcal{Y} \mid \sigma\{E(f \mid \mathcal{A}) : f \in \mathcal{Y}^+\}$. Hence, $\mathcal{B}_{\text{min}} \subset \sigma\{E(f \mid \mathcal{A}) : f \in \mathcal{Y}^+\}$.

Thus, the minimal sufficient parameter $\mathcal{B}_{\text{min}}$ is equal to $\sigma\{E(f \mid \mathcal{A}) : f \in \mathcal{Y}^+\}$. Therefore, b-identifiability is a genuinely Bayesian concept since it depends on the prior distribution through the prior-null sets. For details and properties, see Florens *et al.* (1990, chapter 4).

Acknowledgments The first author gratefully acknowledges the financial support of the FONDECYT Project N$^{\text{o}}$ 1060722 from the Chilean Government. The second author gratefully acknowledges the financial support of the FNRS from the Belgian Government to visit the Department of Statistics, Pontificia Universidad Católica de Chile, Chile.

References

Adams, R. J., Wilson, M. and Wang, W. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement* **21**, 1–23.

Andersen, E. B. (1973). *Conditional Inference and Models for Measuring*. Mentalhygiejnisk Forlag, Copenhagen.

Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. North-Holland, Amsterdam.

Andersen, E. B. and Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psyhometrika* **42**, 357–374.

Baker, F. B. and Kim, S. -H. (2004). *Item Response Theory. Parameter Estimation Techniques. Second Edition, Revised and Expanded*. Marcel and Dekker, New York.

Barankin, E. W. (1961). Sufficient parameters: solution to the minimal dimensionality problem. *Annals of the Institute of Statistical Mathematics* **12**, 91-118.

Barankin, E. W., Kudo, H. and Kusama, T. (1980). Specification of Statistical Models by Sufficiency. In: K. Matusita (Ed.), *Recent Developments in Statistical Inference and Data Analysis*, pp. 9-19. North-Holland, Amsterdam.

Bechger, T. M., Maris, G., Verstralen, H. H. F. M. and Béguin, A. A. (2003). Using Classical Test Theory in Combination With Item Response Theory. *Applied Psychological Measurement* **27**, 319–334.

Bock, R.D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46**, 443–459; **47**, 369 (Errata).

Caillot, P. and Martin, F. (1972). Le modèle bayésien. *Annales de l'Institut Henri Poincaré, section B* **8**, 183–210.

Chen, M. -H. and Dey, D. K. (1998). Bayesian Modeling of Correlated Binary Responses via Scale Mixture of Multivariate Normal Link Functions. *Sankhya: The Indian Journal of Statistics* **60**, 323–343.

Clifford, P. (1982). Some General Comments on Nonidentifiability. In: L. Le Cam and J. Neyman (Eds.), *Probability Models and Cancer*, pp. 81-83. North Holland, Amsterdam.

Clifford, P. (1982a). The Limits of Nonidentifiability in Time-dependent Comportment Models with Applications to Serial-sacrifice Experiments. In: L. Le Cam and J. Neyman (Eds.), *Probability Models and Cancer*, pp. 85-97. North Holland, Amsterdam.

De Boeck, P. and Wilson, M. (2004). *Explanatory Item Response Models. A Generalized Linear and Nonlinear Approach*. Springer, New York.

De Leeuw, J. and Verhelst, N. (1986). Maximum Likelihood Estimation in Generalized Rasch Models. *Journal of the Educational Statistics* **11**, 183–196.

Ellis, J. L. and Junker, B. W. (1997). Tail-Measurability in Monotone Latent Variables Models. *Psychometrika* **62**, 495–523.

Fischer, G. H. (1995). Derivations of the Rasch Model. In: G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models. Foundations, Recent Developments and Applications*, chapter 2. Springer, New York.

Florens, J. -P., Mouchart, M. and Rolin, J. -M. (1985). On Two Definitions of Identification. *Statistics* **16**, 213–218.

Florens, J. -P., Mouchart, M. and Rolin, J. -M. (1990). *Elements of Bayesian Statistics*. Marcel and Dekker, New York.

Florens, J. -P., Mouchart, M. and Rolin, J. -M. (1993). Noncausality and marginalization of Markov processes. *Econometric Theory* **9**, 241–262.

Florens, J. P. and Rolin, J. -M. (1984). Asymptotic sufficiency and exact estimability in Bayesian experiments. In: J. P. Florens, M. Mouchart, J. -P. Raoult and L. Simar, *Alternative Approaches to Time Series Analysis*, pp. 121–142. Bruxelles, Publication des Facultés Universitaires Saint Louis.

Gelfand, A. E. and Sahu, S. K. (1999) Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association* **94**, 247-253.

Ghosh, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statistical and Probability Letters* **23**, 165-170.

Ghosh, M., Ghosh, A., Chen, M. -H. and Agresti, A. (2000). Noninformaive priors for one-parameter item response models. *Journal of Statistical Planning and Inference* **88**, 99–115.

Gustafson, P. (2005). On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables (with Discussion). *Statistical Science* **20**, 111-.140.

Hurwicz, L. (1950). Generalization of the Concept of Identification. In: T. C. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models*, pp. 245-257. Wiley, New York.

Hutchinson, T. P. (1991). *Ability, Partial Information. Guessing: Statistical Modelling Applied to Multiple-Choice Tests*. Rumsby Scientific Publishing, Adelaide, South Australia.

Kadane, J. (1974). The role of identification in Bayesian theory. In: S. Fienberg and A. Zellner (Eds.), *Studies in Bayesian Econometrics and Statistics*, pp. 175-191. North-Holland, Amsterdam.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *The Annals of Mathematical Statistics* **27**, 887–906.

Koopmans, T. C. and Reiersøl, O. (1950). The Identification of Structural Characteristics. *The Annals of Mathematical Statistics* **21**, 165–181.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics* **95**, 391–413.

Leamer, E. E. (1978) *Specification Searches. Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.

Lindley, D. V. (1971). *Bayesian Statistics, A Review*. Society for Industrial and Applied Mathematics, Philadelphia.

Lord, F. M. (1952). *A Theory of Test Scores*. Psychometric Monograph Number 7.

Lord, F. M. (1983). Maximum Likelihood Estimation of Item Response Parameters When Some Responses Are Omitted. *Psychometrika* **48**, 477–482.

Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Harvard University Press, London.

Molenaar, I. W. (1995). Estimation of Item Parameters. In: G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models. Foundations, Recent Developments and Applications*, chapter 3. Springer, New York.

Mellenbergh, G. (1994). Generalized linear item response theory. *Psychological Bulletin* **115**, 300–307.

Mislevy, R. J. and Bock, R. D. (1982). Biweight estimates of latent ability. *Educational and psychological measurement* **42**, 725–737.

Neveu, J. (1964). *Bases Mathématiques du Calcul des Probabilités*. Masson, Paris.

Neyman, J. and Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16**, 1-32.

Oulhaj, A. and Mouchart, M. (2003). Partial sufficiency with connection to the identification problem. *METRON - International Journal of Statistics* **61**, 267–283.

Pfanzagl, J. (1970). Consistent Estimation in the Presence of Incidental Parameters. *Metrika* **15**, 141–148.

Picci, G. (1977). Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem. *SIAM Journal of Applied Mathematics* **33**, 383–398.

Poirier, D. J. (1998) Revising Beliefs in Nonidentified Models. *Econometric Theory* **14**, 483-509.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. The Danish Institute for Educational Research, Copenhagen (Expanded Edition, 1980, The University Chicago Press, Chicago).

Rasch, G. (1966). An Item Analysis which takes Individual Differences into Account. *British Journal of Mathematical and Statistical Psychology* **19**, 49-57.

Rijmen, F. and De Boeck, P. (2005). A Relation Between a Between-Item Multidimensional IRT Model and the Mixture Rasch Model. *Psychometrika* **70**, 481–496.

Rijmen, F., Tuerlinkx, F., De Boeck, P. and Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods* **8**, 185–205.

San Martín, E., Mouchart, M. and Rolin, J. -M. (2005). Ignorable Common Information, Null Sets and Basu's First Theorem. *Sankhya: The Indian Journal of Statistics* **67**, 674-698.

Sijtsma, K. and Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response Theory*. Sage Publications, Thousand Oaks.

Smits, D. J. M. and Moore, S. (2004). Latent item predictors with fixed effects. In: P. De Boeck and M. Wilson (Eds.), *Explanatory Item Response Models. A Generalized Liner and Nonlinear Approach*, chapter 9. Springer, New York.

Swartz, T., Haitovsky, Y., Vexler, A. and Yang, T. (2004). Bayesian identifiability and misclassification in multinomial data. *Canadian Journal of Statistics* **32**, 1–18.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* **47**, 175–186.

Tuerlinckx, F., Rijmen, F., Verbeke, G. and De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology* **59**, 225–255.