PSYCHOMETRIKA—VOL. 78, NO. 2, 341–379
APRIL 2013
DOI: 10.1007/S11336-013-9322-8

# IDENTIFICATION OF THE 1PL MODEL WITH GUESSING PARAMETER: PARAMETRIC AND SEMI-PARAMETRIC RESULTS

ERNESTO SAN MARTÍN

FACULTY OF MATHEMATICS, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTY OF EDUCATION, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

MEASUREMENT CENTER MIDE UC

CEPPE

JEAN-MARIE ROLIN

INSTITUT DE STATISTIQUE, BIOSTATISTIQUE ET SCIENCES ACTUARIELLES, UNIVERSITÉ CATHOLIQUE DE LOUVAIN

LUIS M. CASTRO

DEPARTMENT OF STATISTICS, UNIVERSIDAD DE CONCEPCIÓN

In this paper, we study the identification of a particular case of the 3PL model, namely when the discrimination parameters are all constant and equal to 1. We term this model, 1PL-G model. The identification analysis is performed under three different specifications. The first specification considers the abilities as unknown parameters. It is proved that the item parameters and the abilities are identified if a difficulty parameter and a guessing parameter are fixed at zero. The second specification assumes that the abilities are mutually independent and identically distributed according to a distribution known up to the scale parameter. It is shown that the item parameters and the scale parameter are identified if a guessing parameter is fixed at zero. The third specification corresponds to a semi-parametric 1PL-G model, where the distribution $G$ generating the abilities is a parameter of interest. It is not only shown that, after fixing a difficulty parameter and a guessing parameter at zero, the item parameters are identified, but also that under those restrictions the distribution $G$ is not identified. It is finally shown that, after introducing two identification restrictions, either on the distribution $G$ or on the item parameters, the distribution $G$ and the item parameters are identified provided an infinite quantity of items is available.

Key words: 2PL model, 3PL model, location-scale distributions, fixed effects, random effects, identified parameter, parameters of interest, Hilbert space, Gibbs sampler, measurable separability.

## 1. Introduction

For multiple-choice tests, it is reasonable to assume that the respondents guess when they believe that they do not know the correct response. This type of behavior seems to be prevalent in a low-stakes test, where students are asked to take a test for which they receive neither grades nor academic credit, and thus may be unmotivated to do well. A solution for this problem is to combine the 1PL or 2PL model with a so-called guessing parameter. Three possibilities are found in the literature (Hutschinson, 1991): (1) a fixed value $L^{-1}$, with $L$ being the number of response categories; (2) an overall guessing parameter to be estimated from the data, with the same value for all items; (3) an item-specific guessing parameter. The second possibility can be

Requests for reprints should be sent to Ernesto San Martín, Faculty of Mathematics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile. E-mail: esanmart@mat.puc.cl

viewed as a restricted 3PL in the sense that the guessing parameters are all equal to a single unknown value. The third possibility is the one used in the 3PL, the extension of the 2PL with an item-specific guessing parameter; this parameter reflects the probability of a correct guess. When the discrimination parameters are assumed to be 1, another restricted 3PL is obtained, the so-called 1PL-G model according to San Martín, Del Pino, and De Boeck's (2006) terminology.

The 3PL, and its restricted versions, are popular models. It is not only discussed in most handbooks of IRT (Hambleton, Swaminathan, & Rogers, 1991; van der Linden & Hambleton, 1997; McDonald, 1999; Embretson & Reise, 2000; Thissen & Wainer, 2001; Millsap & Maydeu-Olivares, 2009), but the option is available in various specialized computer programs (e.g., BILOG, LOGIST, MIRTE, MULILOG, PARSCALE, RASCAL, ltm). In spite of being widely used, there are still basic questions to be investigated regarding the 3PL model, and these questions are likely to have an impact on both theory and practice. Parameter identification is one of those basic questions. Identifiability is relevant because it is a condition necessary for ensuring a coherent inference on the parameters of interest. The parameters of interest are related to the sampling distributions, which describe the data generating process. If a one-to-one relationship does not exist between those parameters and the sampling distributions, the parameters of interest are not provided with an empirical meaning. In a sampling theory framework, this fact is made explicit through the impossibility of obtaining unbiased and/or consistent estimators of unidentified parameters (Koopmans & Reiersøl, 1950; Gabrielsen, 1978; San Martín & Quintana 2002). This limitation seems to be circumvented by a Bayesian approach because in this set-up it is always possible to compute the posterior distribution of unidentified parameters (Lindley, 1971; Poirier, 1998; Gelfand & Sahu 1999; Ghosh, Ghosh, Chen, & Agresti, 2000; Gustafson, 2005). However, taking into account that a statistical model always involves an identified parametrization (see Florens, Mouchart, & Rolin, 1990, Theorem 4.3.3), it can be shown that the posterior distribution of an unidentified parameter updates the identified parameter only, and consequently, does not provide any empirical information on the unidentified parameter (San Martín & González, 2010; San Martín, Jara, Rolin, & Mouchart, 2011). Thus, either from a Bayesian point of view or from a sampling theory framework, identification is an issue that needs to be considered.

This paper intends to contribute to the complex problem of identifying the 3PL model. It considers the identification of one of the restricted 3PL models mentioned above, namely the 1PL-G model. As is well known, an identification analysis should begin by making explicit the sampling distributions as well as the parameters of interest. Three different types of likelihoods can be considered:

1. A first likelihood corresponds to the probability distribution of the observations, and it is indexed by both the item parameters (difficulty and guessing) and the abilities.
2. A second likelihood is obtained after integrating out the abilities, which in turn are assumed to be distributed according to a parametric distribution $G^{\boldsymbol{\varphi}}$ known up to a parameter $\boldsymbol{\varphi}$. The sampling distribution or likelihood is accordingly indexed by the item parameters (difficulty and guessing) and $\boldsymbol{\varphi}$. The abilities are typically obtained in a second step through an empirical Bayes procedure.
3. A third likelihood is obtained after integrating out the abilities, which in turn are assumed to be distributed according to an unknown probability distribution $G$. The sampling distribution or likelihood is accordingly indexed by the item parameters (difficulty and guessing) and $G$.

In each of these cases, the parameters indexing the likelihoods represent the parameters of interest. Let us mention that, for the 3PL model, these three perspectives are found in the psychometric literature. For the first type of likelihood, see Swaminathan and Gifford (1986) and Maris and Bechger (2009); for the second type, see Bock and Aitkin (1981) and Bock and Zimowski (1997);

for the third type, see Woods (2006, 2008). Following the terminology of generalized (non-)linear mixed models (De Boeck & Wilson, 2004), it can be said that the first likelihood considers the abilities as *fixed-effects*, whereas the remaining two likelihoods consider the abilities as *random-effects*. This terminology helps to make precise whether the abilities are parameters indexing or not the likelihood function. It can be used in spite of the estimation procedure, particularly in a Bayesian framework, where the parameters indexing the likelihood are endowed with a prior distribution. For identification purpose, the estimation procedure is irrelevant and, therefore, this terminology seems to be useful.

The identification problems corresponding to each of those sampling distributions are quite different. Some contributions conjecture (Adams, Wilson, & Wang, 1997; Adams & Wu, 2007) that the identification of the parameters indexing the first type of likelihoods implies the identification of the parameters indexing the second (and, by extension, the third) type of likelihoods. However, in the case of the Rasch or 1PL model, it has been shown that such relationships are not true (see San Martín et al., 2011, Sections 3.2 and 4.2). This result suggests that the identification problems in the context of 1PL-G models are still open problems.

This paper focuses its attention on these identification problems for the 1PL-G model. It is mainly motivated by the recent contribution of Maris and Bechger (2009); there it is considered a likelihood of the first type, where the discrimination parameters are equal to an unknown common value. They showed that the item parameters and the abilities are not identified. Our paper studies the identification problem in the three contexts mentioned above: Section 2 discusses the problem under a fixed-effects specification of the 1PL-G model; Section 3 studies the problem under a parametric random-effects specification, where the distribution generating the abilities is assumed to be known up to a scale parameter and a location parameter. Finally, in Section 4, the problem is analyzed in a semi-parametric context, where the distribution generating the abilities is considered among the parameters of interest. The main results can be summarized as follows:

1. Under a fixed-effects specification of the 1PL-G model, the item parameters and the abilities are identified if basically one difficulty and one guessing are fixed at 0.
2. Under a parametric random-effects specification, the item parameters and the scale parameter are identified if basically one guessing is fixed at 0.
3. Under a semi-parametric specification, the item parameters are identified if basically one difficulty and one guessing are fixed at 0. Using the structure of a specific Hilbert space, it is shown that the distribution $G$ is not identified by the observations. However, when an infinite quantity of items is available, it is proved that $G$ becomes identified.

These results are relevant because they show under which conditions a specific guessing behavior, captured by an item-specific guessing parameter, has an empirical sense. Furthermore, these results suggest how the identification of the 3PL could rigorously be obtained.

In Section 5, two additional topics are discussed in a general way. First, why the updating of unidentified parameters reduces to the updating of identified parameters. Second, what is the possible impact of unidentifiability on Bayesian estimation procedures. The paper concludes with a discussion, where the practical consequences of our identification results are summarized.

## 2. Identification Under a Fixed-Effects Specification

The 3PL model, introduced by Birnbaum (1968), is specified as follows: for each person $i = 1, \ldots, N$ and each item $j = 1, \ldots, J$, the probability that person $i$ answers correctly item $j$ is given by

$$P[Y_{ij} = 1 \mid \theta_i, \beta_j, \alpha_j, c_j] = c_j + (1 - c_j)\Psi\big[\alpha_j(\theta_i - \beta_j)\big], \tag{2.1}$$

where $\Psi(x) = \exp(x)/[1+\exp(x)]$. This model assumes that if a person $i$ has ability $\theta_i$, then the probability that he/she will know the correct answer of the item $j$ is given by $\Psi[\alpha_j(\theta_i - \beta_j)]$; here $\alpha_j$ corresponds to the discrimination parameter of item $j$, whereas $\beta_j$ is the corresponding difficulty parameter. It further assumes that if he/she does not know the correct answer, he/she will guess and, with probability $c_j$, will guess correctly. The parameter $c_j$ is accordingly called the guessing parameter of item $j$. It follows from these assumptions that the probability of a correct response to item $j$ by person $i$ is given by (2.1). For details, see Birnbaum (1968) and Embretson and Reise (2000, Chapter 4).

The model is completed by assuming that the $\{Y_{ij} : i = 1, \dots, N; j = 1, \dots, J\}$ are mutually independent. The statistical model (or, likelihood function) describing the data generating process corresponds, therefore, to the family of sampling distributions indexed by the parameters

$$(\boldsymbol{\theta}_{1:N}, \boldsymbol{\alpha}_{1:J}, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}) \in \mathbb{R}^N \times \mathbb{R}_+^J \times \mathbb{R}^J \times [0,1]^J,$$

where $\boldsymbol{\theta}_{1:N} = (\theta_1, \dots, \theta_N)$, $\boldsymbol{\alpha}_{1:J} = (\alpha_1, \dots, \alpha_J)$, and similarly for $\boldsymbol{\beta}_{1:J}$ and $\boldsymbol{c}_{1:J}$.

Recently, Maris and Bechger (2009) considered the identification of a particular case of the 3PL, namely when the discrimination parameters $\alpha_j$, with $j = 1, \dots, J$, are equal to $\alpha$. The parameter indeterminacies inherited from the Rasch model and the 2PL model that should be removed, are the location and scale ones. Maris and Bechger (2009) removed them by fixing $\alpha$ at one and constraining $\beta_1$ and $c_1$ in such a way that $\beta_1 = -\ln(1 - c_1)$. In this specific case, the unidentifiability of the parameters of interest persists because $(\theta_i, \beta_j, c_j)$ and $(\ln(\exp(\theta_i) + r)$, $\ln(\exp(\beta_j) - r)$, $(c_j \exp(\beta_j) - r)/(\exp(\beta_j) - r))$, with a constant $r$ such that

$$-\min\bigl\{\exp(\theta_i) : i = 1, \dots, N\bigr\} \le r \le \min\bigl\{c_j \exp(\beta_j) : j = 1, \dots, J\bigr\},$$

induce the same probability distribution (2.1) (with $\alpha_j = 1$ for all item $j$). Thus, "in contrast to the location and scale indeterminacy, this new form of indeterminacy involves not only the ability and the item difficulty parameters, but also the guessing parameter" (Maris & Bechger 2009, p. 6).

The question is under which additional restrictions, the parameters $(\boldsymbol{\theta}_{1:N}, \alpha, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J})$ are identified by the observations. San Martín, González, and Tuerlinckx (2009) considered this problem when $\alpha = 1$. This is the case of the 1PL-G model, which is specified as

$$P[Y_{ij} = 1 \mid \theta_i, \beta_j, c_j] = c_j + (1 - c_j)\Psi(\theta_i - \beta_j). \tag{2.2}$$

The data generating process is accordingly described by a family of sampling distributions indexed by the parameters

$$(\boldsymbol{\theta}_{1:N}, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}) \in \mathbb{R}^N \times \mathbb{R}^J \times [0,1]^J. \tag{2.3}$$

San Martín et al. (2009) shown that these parameters are identified by the observations under specific restrictions which are summarized in the following theorem:

**Theorem 1.** *For the statistical model* (2.2), *the parameters* $(\boldsymbol{\theta}_{1:N}, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J})$ *are identified by the observations provided the following conditions hold*:

1. *There exists at least two persons such that their probabilities to correctly answer all the items are different*.
2. *The parameters* $c_1$ *and* $\beta_1$ *are fixed at* 0.

This theorem deserves the following comments:

1. Under the constraints of Theorem 1, the person specific parameter $\theta_i$ is equal to $\ln\{P[Y_{i1}=1 \mid \theta_i]/P[Y_{i1}=0 \mid \theta_i]\}$, the betting odds of a correct answer of person $i$ to the standard item 1. Thus, under the assumption that there exists an item (the standard item whose difficulty is fixed at 0) that persons can not answer it correctly by guessing, abilities can be compared through odd ratios in the same way as it is done in the context of a fixed-effects 1PL model.
2. Theorem 1 is still valid if the difficulty parameters $\beta_j$'s are equal to a common value $\beta$ and the guessing parameters $c_j$'s are equal to a common value $c$. In this case, only the person-specific parameters would be identified.
3. In practice, the ability parameters $\boldsymbol{\theta}_{1:N}$ are used to decide who passes or fails a test, used to decide which students should receive more advance or remedial instruction, etc. If this procedure is done using the 1PL-G model, Theorem 1 provides identification restrictions under which not only the item parameters are identified, but also the person specific parameters. These parameters can, for instance, be estimated using either a joint maximum likelihood (JML) estimator, or a Bayesian procedure. Although JML is in principle feasible, it is known that it produces biased estimations for the 1PL model due to the incidental parameter problem (Andersen, 1980; Lancaster, 2000; Del Pino, San Martín, González, & de Boeck, 2008). A solution to this problem is to estimate IRT models using a marginal maximum likelihood (MML) estimator (Molenaar, 1995; Thissen, 2009), where the person specific abilities are assumed to be distributed according to a distribution $G^{\boldsymbol{\varphi}}$ known up to a parameter $\boldsymbol{\varphi}$. The person specific ability is estimated through an empirical Bayes procedure. If a Bayesian approach is chosen, a prior distribution on the identified parameters $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J}, \boldsymbol{\theta}_{1:N})$ should be specified. It is often assumed that $\boldsymbol{\beta}_{2:J}$, $\boldsymbol{c}_{2:J}$ and $\boldsymbol{\theta}_{1:N}$ are a priori independent; in some applications, the distribution $G^{\boldsymbol{\varphi}}$ is considered as the prior distribution on $\boldsymbol{\theta}_{1:N}$, causing $\boldsymbol{\varphi}$ to become a hyperparameter; see, for instance, Béguin and Glas (2001) for the 3PNO model.

## 3. Identification of the Parametric 1PL-G Model

### 3.1. *Random-Effects Specification of the 1PL-G Model*

The previous identification results are valid under a *fixed-effects specification* of the 1PL-G model, that is, when the abilities are viewed as *unknown parameters*. However, as mentioned above, in modern item response theory, $\theta$ is usually considered as a latent variable and, therefore, its probability distribution is an essential part of the model. The probability distribution generating the person specific abilities $\theta_i$'s is assumed to be a location-scale distribution $G^{\mu,\sigma}$ defined as

$$P[\theta_i \leq x \mid \mu, \sigma] = G^{\mu,\sigma}\big((-\infty, x]\big) \doteq G\left(\left(-\infty, \frac{x-\mu}{\sigma}\right]\right), \tag{3.1}$$

where $\mu \in \mathbb{R}$ is the location parameter and $\sigma \in \mathbb{R}_+$ is the scale parameter. In applications, $G$ is typically chosen as a standard normal distribution.

It is also assumed that for each person $i$, his/her responses $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$ satisfy the Axiom of Local Independence, namely that $Y_{i1}, \dots, Y_{iJ}$ are mutually independent conditionally on $(\theta_i, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J})$. The distribution of $Y_{ij}$ depends on $(\theta_i, \beta_j, c_j)$ through the function (2.2). It is finally assumed that, conditionally on $(\boldsymbol{\theta}_{1:N}, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J})$, the response patterns $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N$ are mutually independent.

The statistical model (or, likelihood function) is obtained after integrating out the random effects $\theta_i$'s. The above-mentioned hypotheses underlying the 1PL-G model imply that the response patterns $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N$ are mutually independent given $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma)$, with a common probability function defined as

$$\begin{aligned} &P[\boldsymbol{Y}_i = \boldsymbol{y}_i \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma] \\ &\quad = \int_{\mathbb{R}} \prod_{j=1}^{J} \{P[Y_{ij} = 1 \mid \theta, \beta_j, c_j]\}^{y_{ij}} \{P[Y_{ij} = 0 \mid \theta, \beta_j, c_j]\}^{1-y_{ij}} G^{\mu,\sigma}(d\theta), \end{aligned} \tag{3.2}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iJ})' \in \{0, 1\}^J$ and $P[Y_{ij} = 1 \mid \theta, \beta_j, c_j]$ as defined by (2.2). The parameters of interest are accordingly given by

$$(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma) \in \mathbb{R}^J \times [0, 1]^J \times \mathbb{R} \times \mathbb{R}_+. \tag{3.3}$$

In order to distinguish the probability distribution (3.2) from the *conditional* probability $P[Y_{ij} = 1 \mid \theta_i, \beta_j, c_j]$, (3.2) is termed a *marginal* probability.

Under the *iid* property of the statistical model generating the $\boldsymbol{Y}_i$'s, the identification of the parameters of interest by one observation is entirely similar to their identification by an infinite quantity of observations (for a proof, see Florens et al., 1990, Theorem 7.6.6). Thus, the identification problem to be studied in this section consists of establishing restrictions (if necessary) under which the mapping $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma) \longmapsto P[\boldsymbol{Y}_1 = \boldsymbol{y}_1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma]$ is injective for all $\boldsymbol{y}_1 \in \{0, 1\}^J$, where $P[\boldsymbol{Y}_1 = \boldsymbol{y}_1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma]$ is given by (3.2).

### 3.2. *Identification Strategy*

Following San Martín et al. (2009), an identification strategy consists of distinguishing between parameters of interest and identified parameters. In the case of the statistical model (3.2), the parameters of interest are $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma)$. The probabilities of the $2^J$ different possible patterns are given by

$$\begin{aligned} q_{12\cdots J} &= P[Y_{11} = 1, \ldots, Y_{1,J-1} = 1, Y_{1J} = 1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma] \\ q_{12\cdots \bar{J}} &= P[Y_{11} = 1, \ldots, Y_{1,J-1} = 1, Y_{1J} = 0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma] \\ &\vdots \\ q_{\bar{1}\bar{2}\cdots \bar{J}} &= P[Y_{11} = 0, \ldots, Y_{1,J-1} = 0, Y_{1J} = 0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma]. \end{aligned} \tag{3.4}$$

The statistical model (3.2) corresponds, therefore, to a multi-nomial distribution $(\boldsymbol{Y}_1 \mid \boldsymbol{q}) \sim \text{Mult}(2^J, \boldsymbol{q})$, where $\boldsymbol{q} = (q_{12\cdots J}, q_{12\cdots J-1,\bar{J}}, \ldots, q_{\bar{1},\bar{2},\ldots,\bar{J}})$. It is known that the parameter $\boldsymbol{q}$ of a multi-nomial distribution is identified by $\boldsymbol{Y}_1$. We accordingly term the parameter $\boldsymbol{q}$, *identified parameter*; the $q$'s with less than $J$ subscripts are linear combinations of them and, therefore, are identified by the observations. Consequently, the identification of the parameters of interest follows if a bijective relationship between them and functions of the identified parameters $\boldsymbol{q}$ is established. This strategy is followed in the rest of this paper. Taking into account that a statistical model always involves an identified parameterization (see Florens et al., 1990; San Martín et al., 2009), the restrictions which are introduced to establishing a bijective relationship between the parameters of interest and the identified parameters, are not only sufficient identification conditions, but also necessary conditions.

### 3.3. *Identification when G is Known up to a Scale Parameter*

Let us begin the identification analysis of the 1PL-G model when the distribution $G$ generating the person specific abilities is known up to the scale parameter $\sigma$. The corresponding identification analysis is divided into three steps:

STEP 1: It is shown that the difficulty parameters $\boldsymbol{\beta}_{1:J}$ are a function of the scale parameter $\sigma$, the guessing parameters $\boldsymbol{c}_{1:J}$ and the identified parameters $P[Y_{ij}=0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma]$, with $j = 1, \dots, J$.

STEP 2: It is shown that the scale parameter $\sigma$ is a function of the guessing parameters $c_1$ and $c_2$, as well as of the identified parameters $P[Y_{i1}=1, Y_{i2}=1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma]$, $P[Y_{i1}=1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma]$ and $P[Y_{i2}=1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma]$.

STEP 3: It is finally shown that the difficulty parameters $\boldsymbol{\beta}_{2:J}$ and the guessing parameters $\boldsymbol{c}_{2:J}$ are functions of $(\beta_1, c_1)$ and identified parameters which in turn depend on $\boldsymbol{q}$.

Combining these steps, the identification of $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma)$ by both the observations and $c_1$ is obtained. Therefore, one identification restriction is needed, namely $c_1 = 0$. Let us mention that both STEP 1 and STEP 2 are valid for all the conditional specifications of the form

$$P[Y_{ij}=1 \mid \theta_i, \beta_j, c_j] = c_j + (1-c_j)F(\theta_i - \beta_j), \tag{3.5}$$

where $F$ is a strictly increasing continuous distribution function with a positive density function on $\mathbb{R}$, and not only for the logistic distribution $\Psi$. However, STEP 3 depends on the logistic function $\Psi$. In what follows, these steps are duly detailed.

*3.3.1. Step 1 for the Parametric 1PL-G Model* Let

$$\omega_j \doteq P[Y_{ij}=0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma] = \delta_j \, p(\sigma, \beta_j), \tag{3.6}$$

where

$$p(\sigma, \beta_j) = \int_{\mathbb{R}} \big\{1 - F(\sigma\theta - \beta_j)\big\} G(d\theta), \tag{3.7}$$

with $F$ a strictly increasing continuous distribution function, and $\delta_j \doteq 1 - c_j$. The parameter $\omega_j$ is identified because it is a function of the identified parameter $\boldsymbol{q}$; see Section 3.2.

The function $p(\sigma, \beta_j)$ is a continuous function in $(\sigma, \beta_j) \in \mathbb{R}_+ \times \mathbb{R}$ that is strictly increasing in $\beta \in \mathbb{R}$ because $F$ is a strictly increasing continuous function (in particular, the logistic distribution $\Psi$ satisfies these properties). Furthermore, $p(\sigma, -\infty) = 0$ and $p(\sigma, +\infty) = 1$, and consequently,

$$0 \le \omega_j \le \delta_j \quad \text{for all } j = 1, \dots, J. \tag{3.8}$$

Therefore, if we define

$$\bar{p}(\sigma, \epsilon) \doteq \inf\big\{\beta : p(\sigma, \beta) > \epsilon\big\}, \tag{3.9}$$

it is clear that

$$\bar{p}\big[\sigma, p(\sigma, \beta)\big] = \beta \quad \text{for all } \beta. \tag{3.10}$$

Taking into account relation (3.8), and assuming that $\delta_j > 0$ for each item $j$, it follows that $\beta_j = \bar{p}[\sigma, \omega_j/\delta_j]$; that is, for all $j = 1, \dots, J$, the item parameter $\beta_j$ is a function of the scale parameter $\sigma$, of the identified parameter $\omega_j$ and of the non-guessing parameter $\delta_j$ (and, by extension, of the guessing parameter $c_j$).

*3.3.2. Step 2 for the Parametric 1PL-G Model* Let

$$\begin{aligned}\omega_{12} &\doteq P[Y_{i1}=0, Y_{i2}=0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma] \\ &= \delta_1\delta_2 \int_{\mathbb{R}} \big\{1 - F(\sigma\theta - \beta_1)\big\}\big\{1 - F(\sigma\theta - \beta_2)\big\} G(d\theta).\end{aligned} \tag{3.11}$$

Using STEP 1, the identified parameter $\omega_{12}$ can be written as a function of $\sigma, \delta_1, \delta_2, \omega_1$, and $\omega_2$ in the following way:

$$\omega_{12} \doteq \varphi(\sigma, \delta_1, \delta_2, \omega_1, \omega_2)$$
$$= \delta_1\delta_2 \int_{\mathbb{R}} \left\{1 - F\left[\sigma\theta - \overline{p}\left(\sigma, \frac{\omega_1}{\delta_1}\right)\right]\right\}\left\{1 - F\left[\sigma\theta - \overline{p}\left(\sigma, \frac{\omega_2}{\delta_2}\right)\right]\right\} G(d\theta).$$

Now, if the distribution function $F$ has a continuous density function $f$ strictly positive on $\mathbb{R}$, then it can be shown that the function $\omega_{12} = \varphi(\sigma, \delta_1, \delta_2, \omega_1, \omega_2)$ is a strictly increasing continuous function of $\sigma$ and, therefore, $\sigma = \overline{\varphi}(\omega_{12}, \delta_1, \delta_2, \omega_1, \omega_2)$, where

$$\overline{\varphi}(\omega, \delta_1, \delta_2, \omega_1, \omega_2) = \inf\left\{\sigma : \varphi(\sigma, \delta_1, \delta_2, \omega_1, \omega_2) > \omega\right\}. \quad (3.12)$$

In other words, $\sigma$ becomes a function of the identified parameters $\omega_1, \omega_2$ and $\omega_{12}$, as well as of the non-guessing parameters $\delta_1$ and $\delta_2$. The details are developed in Appendix A.

*3.3.3. Step 3 for the Parametric 1PL-G Model* This step essentially depends on the logistic distribution $\Psi$, and consequently, the arguments below are performed using the conditional probability (2.2). Let $J \geq 3$, $j \neq 1$ and $k \neq j$ (with $k, j \leq J$). Define the identified parameters $p_0^J$, $p_j^J$ and $p_{jk}^J$ as follows:

$$p_0^J \doteq P\left[\bigcap_{1\leq j\leq J}\{Y_{ij} = 0\} \,\Big|\, \boldsymbol{\beta}_{1:J}, \boldsymbol{\delta}_{1:J}, \sigma\right] = \prod_{1\leq j\leq J} \delta_j \times I_0^J(\boldsymbol{\beta}_{1:J}, \sigma);$$

$$p_j^J \doteq P\left[\bigcap_{1\leq k\leq J, k\neq j}\{Y_{ik} = 0\} \,\Big|\, \boldsymbol{\beta}_{1:J}, \boldsymbol{\delta}_{1:J}, \sigma\right] = \prod_{1\leq k\leq J, k\neq j} \delta_k\left\{I_0^J(\boldsymbol{\beta}_{1:J}, \sigma) + e^{-\beta_j} I_1^J(\boldsymbol{\beta}_{1:J}, \sigma)\right\};$$

$$p_{jk}^J \doteq P\left[\bigcap_{1\leq r\leq J, r\neq j, r\neq k}\{Y_{ir} = 0\} \,\Big|\, \boldsymbol{\beta}_{1:J}, \boldsymbol{\delta}_{1:J}, \sigma\right]$$
$$= \prod_{1\leq r\leq J, r\neq j, r\neq k} \delta_r\left\{I_0^J(\boldsymbol{\beta}_{1:J}, \sigma) + \left(e^{-\beta_j} + e^{-\beta_k}\right) I_1^J(\boldsymbol{\beta}_{1:J}, \sigma) + e^{-\beta_j} e^{-\beta_k} I_2^J(\boldsymbol{\beta}_{1:J}, \sigma)\right\},$$

where

$$I_0^J(\boldsymbol{\beta}_{1:J}, \sigma) = \int_{\mathbb{R}} \prod_{1\leq j\leq J} \frac{1}{1 + e^{\sigma\theta - \beta_j}} G(d\theta),$$

$$I_1^J(\boldsymbol{\beta}_{1:J}, \sigma) = \int_{\mathbb{R}} e^{\sigma\theta} \prod_{1\leq j\leq J} \frac{1}{1 + e^{\sigma\theta - \beta_j}} G(d\theta),$$

$$I_2^J(\boldsymbol{\beta}_{1:J}, \sigma) = \int_{\mathbb{R}} e^{2\sigma\theta} \prod_{1\leq j\leq J} \frac{1}{1 + e^{\sigma\theta - \beta_j}} G(d\theta).$$

Taking into account that, for each $j = 1, \ldots, J$, $\delta_j > 0$ and $\beta_j \in \mathbb{R}$, it follows that

$$\frac{p_j^J}{p_0^J} = \frac{1}{\delta_j} + \frac{e^{-\beta_j}}{\delta_j} g(\boldsymbol{\beta}_{1:J}, \sigma),$$

$$\frac{p_{jk}^J}{p_0^J} = \frac{1}{\delta_j\delta_k}\left\{1 + \left(e^{-\beta_j} + e^{-\beta_k}\right) g(\boldsymbol{\beta}_{1:J}, \sigma) + e^{-\beta_j} e^{-\beta_k} h(\boldsymbol{\beta}_{1:J}, \sigma)\right\},$$

where

$$g(\boldsymbol{\beta}_{1:J},\sigma)=\frac{I_1^J(\boldsymbol{\beta}_{1:J},\sigma)}{I_0^J(\boldsymbol{\beta}_{1:J},\sigma)}\quad\text{and}\quad h(\boldsymbol{\beta}_{1:J},\sigma)=\frac{I_2^J(\boldsymbol{\beta}_{1:J},\sigma)}{I_0^J(\boldsymbol{\beta}_{1:J},\sigma)}.$$

Using these definitions, the following three propositions are established.

**Proposition 1.** *Let $J\geq 3$ and $j\neq k$. Then*

$$r_{jk}^J\doteq\frac{p_{jk}^J}{p_0^J}-\frac{p_j^J}{p_0^J}\frac{p_k^J}{p_0^J}=\frac{e^{-\beta_j}e^{-\beta_k}}{\delta_j\delta_k}k(\boldsymbol{\beta}_{1:J},\sigma),$$

*where $k(\boldsymbol{\beta}_{1:J},\sigma)=h(\boldsymbol{\beta}_{1:J},\sigma)-g(\boldsymbol{\beta}_{1:J},\sigma)^2>0$.*

The positivity of $k(\boldsymbol{\beta}_{1:J},\sigma)$ follows after noticing that $g(\boldsymbol{\beta}_{1:J},\sigma)=E_{G_{\boldsymbol{\beta}_{1:J},\sigma}}(e^{\sigma\theta})$ and $h(\boldsymbol{\beta}_{1:J},\sigma)=E_{G_{\boldsymbol{\beta}_{1:J},\sigma}}(e^{2\sigma\theta})$, with

$$G_{\boldsymbol{\beta}_{1:J},\sigma}(d\theta)\doteq\frac{1}{I_0^J(\boldsymbol{\beta}_{1:J},\sigma)}\prod_{1\leq j\leq J}\frac{1}{1+e^{\sigma\theta-\beta_j}}G(d\theta).$$

Thus, Proposition 1 ensures that $r_{jk}^J>0$ and, consequently, the following proposition can be established.

**Proposition 2.** *Let $J\geq 3$ and $j\neq k$, with $j\neq 1$. Then*

$$u_j\doteq\frac{r_{1k}^J}{r_{jk}^J}=\frac{\delta_j}{\delta_1}e^{\beta_j-\beta_1},$$

*and consequently, $u_j>0$. Moreover, $\{u_j:2\leq j\leq J\}$ are identified parameters.*

**Proposition 3.** *Let $J\geq 3$ and $j\neq 1$. Then*

$$v_j\doteq\frac{p_j^J}{p_0^J}u_j-\frac{p_1^J}{p_0^J}=\frac{1}{\delta_1}\left(e^{\beta_j-\beta_1}-1\right),$$

*and consequently, $v_j\in\mathbb{R}$. Moreover, $\{v_j:2\leq j\leq J\}$ are identified parameters.*

Propositions 2 and 3 entail the following identities: for $j=2,\dots,J$,

$$\text{(i)}\quad\delta_j=\frac{u_j\delta_1}{v_j\delta_1+1},\qquad\text{(ii)}\quad\beta_j=\beta_1+\ln(v_j\delta_1+1).\tag{3.13}$$

These identities require that, for each $j=2,\dots,J$, $v_j\delta_1+1>0$. By Proposition 3, this inequality is equivalent to $e^{\beta_j-\beta_1}>0$, which is always true. Thus, the inequality $v_j\delta_1+1>0$ does not restrict the sampling process.

*3.3.4. Main Result for the Parametric 1PL-G Model* Using the previous results, the parameters $(\boldsymbol{\beta}_{1:J},\boldsymbol{c}_{2:J},\sigma)$ can be written as a function of $c_1$, in addition to other identified parameters:

1. Equality (3.13.i) implies that, for each $j = 2, \dots, J$, $c_j$ is a function of $c_1$ and $(u_j, v_j)$.
2. STEP 2 and equality (3.13.i) implies that

$$\sigma = \overline{\varphi}\left[\omega_{12}, 1 - c_1, \frac{u_2(1-c_1)}{v_2(1-c_1)+1}, \omega_1, \omega_2\right] \doteq h(1 - c_1, \omega_1, \omega_2, \omega_{12}, u_2, v_2),$$

   where $\overline{\varphi}$ is defined by (3.12).
3. STEP 1, along with the previous conclusion, imply that

$$\beta_1 = \overline{p}\big[h(1 - c_1, \omega_1, \omega_2, \omega_{12}, u_2, v_2), \omega_1/(1-c_1)\big],$$

   where $\overline{p}$ is defined by (3.9).
4. STEP 1 and STEP 2, along with the previous conclusion, imply that, for each $j = 2, \dots, J$,

$$\beta_j = \overline{p}\left[h(1 - c_1, \omega_1, \omega_2, \omega_{12}, u_2, v_2), \frac{\omega_1}{(1-c_1)}\right] + \ln\big[v_j(1-c_1)+1\big].$$

The previous equalities allow to make explicit the identified parametrization of the random-effects 1PL-G model; this parameterization is a function of the guessing parameter $c_1$. Thus, the sampling process depends on the parameters of interest $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma)$ through $c_1$ only. It is actually fully characterized by the following $2^J - 1$ equations: for $\mathcal{K} \subset \{1, \dots, J\}$,

$$\begin{aligned} P&\left[\bigcap_{j\in\mathcal{K}}\{Y_{ij}=0\} \,\middle|\, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma\right] \\ &= \prod_{j\in\mathcal{K}} u_j \times \int_{\mathbb{R}} \frac{G(d\theta)}{\prod_{j\in\mathcal{K}}\{v_j + \frac{1}{1-c_1} + \frac{e^{h(1-c_1,\omega_1,\omega_2,\omega_{12},u_2,v_2)\theta}}{(1-c_1)e^{\overline{p}[h(1-c_1,\omega_1,\omega_2,\omega_{12},u_2,v_2),\omega_1/(1-c_1)]}}\}} \\ &= P\left[\bigcap_{j\in\mathcal{K}}\{Y_{ij}=0\} \,\middle|\, c_1\right]. \end{aligned} \tag{3.14}$$

It should be remarked that the information provided by these $2^J - 1$ marginal probabilities is exactly the same as the information provided by the marginal probabilities $q$'s as defined in (3.4).

Consequently, the parameters of interest $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma)$ becomes identified if the guessing parameter $c_1$ is fixed at a specific value. In principle, this value could be arbitrarily chosen. However, according to inequality (3.8), $c_1 \leq P[Y_{i1} = 1 \mid c_1]$. Since the marginal probability $P[Y_{i1} = 1 \mid c_1]$ is determined empirically, $c_1$ should be fixed at 0. By so doing, the same items can be applied to various samples of persons.

We summarize these findings in the following theorem.

**Theorem 2.** *For the statistical model (3.2) induced by both the 1PL-G model (2.2) and the person specific abilities distributed according to a distribution G known up to the scale parameter $\sigma$, the parameters of interest $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma)$ are identified by $\boldsymbol{Y}_1$ provided that*

1. *At least three items are available.*
2. *The guessing parameter $c_1$ is fixed at* 0.

*Moreover, the specification of the model entails that*

$$0 \leq c_j \leq P[Y_{ij} = 1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma] \quad \textit{for every } j = 2, \ldots, J, \tag{3.15}$$

*where*

$$c_j = 1 - \frac{p_0^J p_{1k}^J - p_1^J p_k^J}{p_{jk}^J(p_0^J - p_1^J) + p_j^J(p_{1k}^J - p_k^J)}, \quad \textit{with } k \neq 1, j. \tag{3.16}$$

This theorem deserves the following comments:

1. Condition (2) of Theorem 2 implies that the marginal probability that a person $i$ correctly answers the item 1 is given (see relations (3.6) and (3.7)) by

$$P[Y_{i1} = 1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma] = \int_{\mathbb{R}} \frac{\exp(\sigma\theta - \beta_1)}{1 + \exp(\sigma\theta - \beta_1)} G(d\theta).$$

   This means that the test (or measurement instrument) must contain an item (labeled with the index 1) such that each person answers it *without* guessing. In other words, the identification restriction $c_1 = 0$ implies a restriction on the design of the multiple-choice test, an aspect that the practitioner should carefully consider. A relevant question is the following: How can it be ensured that each person answers a specific item without guessing? By including in the test a *constructed response item*: each person correctly or incorrectly answer it, but it is not possible to correctly answering it by guessing.[1]
2. Inequality (3.15) not only ensures the possibility to test the null hypothesis $c_j = 0$, but also to specify the probability of a correct answer with the 1PL for some items, and with the 1PL-G for other items. This last situation corresponds to mixed format tests where some items are multiple choice, while some other items are constructed responses, scored right or wrong by some scoring mechanism.
3. According to equality (3.16), the guessing parameter $c_j$ is a function of marginal probabilities, which in turn can be estimated from the data through relative frequencies. This suggests to use these relative frequencies as an estimator of $c_j$. It should be mentioned that these estimates do not depend on the distribution $G^\sigma$ generating the person specific abilities.
4. It should also be remarked that $\beta_j = \beta_1$ for every $j = 2, \ldots, J$ if and only if $v_j = 0$ for every $j = 2, \ldots, J$. In this case, $(\beta_1, \boldsymbol{c}_{2:J}, \sigma)$ are identified provided that $c_1$ is fixed at 0. Similarly, $c_j = c_1$ for every $j = 2, \ldots, J$ if and only if $v_j(1 - c_1) + 1 = u_j$ for every $j = 2, \ldots, J$. This last equality implies that $c_1$ is identified and, therefore, $(\boldsymbol{\beta}_{1:J}, c_1, \sigma)$ are identified *without* identification restrictions.

Let us summarize the last remark in the following corollary:

**Corollary 1.** *Consider the statistical model* (3.2) *induced by both the* 1*PL-G model* (2.2) *and the person specific abilities distributed according to a distribution G known up to the scale parameter* $\sigma$. *If at least three items are available, then the following statements hold*:

1. *If the items have a common difficulty parameter* $\beta_1$, *then* $(\beta_1, \boldsymbol{c}_{2:J}, \sigma)$ *are identified by one observation if* $c_1$ *is fixed at* 0.
2. *If the items have a common guessing parameter* $c_1$, *then* $(\boldsymbol{\beta}_{1:J}, c_1, \sigma)$ *are identified by one observation*.

[1] This suggestion is due to Paul De Boeck.

*3.3.5. Identification when G is Known up to Both a Location and a Scale Parameter* If the distribution $G$ generating the abilities is known up to both a location parameter $\mu$ and a scale parameter $\sigma$, it is then necessary to impose an identification restriction not only on a guessing parameter, but also on the difficulty parameters. As a matter of fact, let $G^{\mu,\sigma}$ be a probability distribution given by (3.1). Relation (3.6) is, therefore, rewritten as

$$\widetilde{\omega_j} \doteq P[Y_{ij} = 0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{\delta}_{1:J}, \sigma, \mu] = \delta_j \int_{\mathbb{R}} \bigl\{1 - F(\sigma\theta + \mu - \beta_j)\bigr\} G(d\theta)$$

for all $j = 1, \ldots, J$. Since $F$ is a strictly increasing continuous function, we have that, for all $j = 1, \ldots, J$, $\beta_j - \mu$ is a function of $(\sigma, \widetilde{\omega_j}, \delta_j)$. Following the arguments developed in Sections 3.3.2 and 3.3.3, it follows that $(\beta_1 - \mu, \ldots, \beta_J - \mu, c_2, \ldots, c_J, \sigma)$ is identified by the observations given $c_1$. Therefore, under a restriction of the form $\boldsymbol{a}'\boldsymbol{\beta}_{1:J} = 0$ such that $\mathbb{1}'_J\boldsymbol{a} \neq 0$, with $\boldsymbol{a} \in \mathbb{R}^J$ known and $\mathbb{1}_J = (1, \ldots, 1)' \in \mathbb{R}^J$, the parameters $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \mu, \sigma)$ are identified by the observations. Summarizing, we establish the following corollary.

**Corollary 2.** *For the statistical model* (3.2) *induced by both the* 1*PL-G model* (2.2) *and the person specific abilities distributed according to a distribution G known up to both the location parameter* $\mu$ *and the scale parameter* $\sigma$, *the parameters of interest* $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \mu, \sigma)$ *are identified by* $\boldsymbol{Y}_1$ *provided that*

1. *At least three items are available.*
2. *The guessing parameter* $c_1$ *is fixed at* 0.
3. $\boldsymbol{a}'\boldsymbol{\beta}_{1:J} = 0$ *such that* $\mathbb{1}'_J\boldsymbol{a} \neq 0$, *with* $\boldsymbol{a} \in \mathbb{R}^J$ *known.*

This corollary deserves the following comments:

1. Condition (3) imposes a restriction on the difficulty parameters. Typical choices are $\boldsymbol{a} = \mathbb{1}_J$, which leads to restricting the difficulty parameters as $\sum_{j=1}^{J} \beta_j = 0$; or $\boldsymbol{a} = (1, 0, \ldots, 0)'$, which leads to imposing $\beta_1 = 0$. It is often said that this type of restrictions lead to fix the scale of the difficulty parameters. However, after considering the proof underlying this corollary, it can be said that, on the one hand, such restrictions lead to separate the location parameter from the difficulty parameters and, on the other hand, excludes equal difficulties of all the items. This is not the case when the person specific abilities are distributed according to $G^\sigma$. In practice this means that, if the data are analyzed using a 1PL-G model with $(\theta_i \mid \mu, \sigma) \sim G^{\mu,\sigma}$, the multiple-choice test is supposed to contain at least two items with different difficulties, in addition to an item such that $c_1 = 0$. This structure of the measurement instrument is assumed by the 1PL-G model when $(\theta_i \mid \mu, \sigma) \sim G^{\mu,\sigma}$, so it can not be statistically tested by the model.
2. The arguments underlying Corollary 2 lead to stating the following remark. Consider $\mu = 0$; Equations (3.6) and (3.7) suggest that the parameter $\sigma$ can be viewed as a discrimination parameter common to all the items. If this is the case, then $G$ should be fully known. If not, let $\alpha$ be the discrimination parameter common to all the items and let $\sigma$ be the scale parameter of $G$. Defining $\widetilde{\sigma} \doteq \sigma\alpha$, the arguments developed in Steps 1, 2, and 3, lead to the conclusion that $\widetilde{\sigma}$ is identified. Therefore, $(\alpha, \sigma)$ and $\widetilde{\sigma}$ are in bijection if either $\alpha$ is fixed at 1, or $\sigma$ is fixed at 1.
3. Let us also remark that these considerations, along with the arguments developed in Steps 1 and 2 above, provide insight to establish the identification of the random-effects 2PL model. Details are provided in Appendix B.

## 4. Identification of the Semi-parametric 1PL-G Model

Many IRT models are fitted under the assumption that $\theta_i$ is normally distributed with an unknown variance. The use of a distribution for $\theta_i$ (as opposed to treating $\theta_i$ as a fixed effect) is a key feature of the widely use marginal maximum likelihood method. The normal distribution is convenient to work with, specially because is available in statistical packages such as SAS (Proc. NLMIXED) or R (lme4, ltm). However, as pointed out by Woods and Thissen (2006) and Woods (2006), there exist specific fields, such as personality and psychopathology, in which the normality assumption is not realistic (for references, see Woods, 2006). In these fields, it could be argued that psychopathology and personality variables are likely to be positively skewed, because most persons in the general population have low pathology, and fewer persons have severe pathology. However, the distribution $G$ of $\theta_i$ is unobservable and, consequently, though a researcher may hypothesize about it, it is not known in advance of an analysis. Therefore, any a priori parametric restriction on the shape of the distribution $G$ could be considered as a mis-specification.

These considerations lead to extending parametric IRT models by considering the distribution $G$ as a parameter of interest and, therefore, to estimating it by using non-parametric techniques. Besides the contributions of Woods and Thissen (2006) and Woods (2006, 2008), Bayesian non-parametric methods applied to IRT models should also be mentioned; see, among others, Roberts and Rosenthal (1998), Karabatsos and Walker (2009), and Miyazaki and Hoshino (2009). In spite of these developments, it is relevant to investigate whether the item parameters as well as the distribution $G$ of an IRT model—in our case the 1PL-G model—are or not identified by the observations. If such parameters are identified, then a semi-parametric extension of the 1PL-G model would provide greater flexibility than does the assumption of non-normal parametric form for $G$.

### *4.1. Semi-parametric Specification of the 1PL-G Model*

A semi-parametric 1PL-G model is obtained after substituting the parametric hypothesis (3.1) by the following hypothesis:

$$(\theta_i \mid G) \overset{\text{iid}}{\sim} G, \tag{4.1}$$

where $G$ is a probability distribution on $(\mathbb{R}, \mathcal{B})$. The rest of the model structure is as specified in Section 3.1. The statistical model is obtained after integrating out the person specific abilities $\theta_i$'s. The response patterns $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$ are mutually independent conditionally on $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$. The common distribution of $\boldsymbol{Y}_i$ is given by

$$P[\boldsymbol{Y}_i = \boldsymbol{y} \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G] = \int_{\mathbb{R}} \prod_{j=1}^{J} \left\{ c_j + (1 - c_j) \frac{\exp[y_j(\theta - \beta_j)]}{1 + \exp(\theta - \beta_j)} \right\} G(d\theta), \tag{4.2}$$

where $\boldsymbol{y} \in \{0, 1\}^J$. The parameters of interest and the corresponding parameter space are

$$(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G) \in \mathbb{R}^J \times [0, 1]^J \times \mathcal{P}(\mathbb{R}, \mathcal{B}).$$

### *4.2. Identification Analysis Under a Finite Quantity of Items*

Similarly to the random-effects 1PL-G model (see Section 3.2), the statistical model induced by the semi-parametric 1PL-G model corresponds to a multi-nomial distribution $\text{Mult}(2^J, \boldsymbol{q})$, where the identified parameter $\boldsymbol{q} = (q_{12\cdots I}, q_{12\cdots I-1,\bar{I}}, \ldots, q_{\bar{1},\bar{2},\ldots,\bar{I}})$ is given by

$$\begin{aligned} q_{12\cdots I} &= P[Y_{11} = 1, \ldots, Y_{1,J-1} = 1, Y_{1J} = 1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G] \\ &\vdots \\ q_{\bar{1}\bar{2}\cdots\bar{I}} &= P[Y_{11} = 0, \ldots, Y_{1,J-1} = 0, Y_{1J} = 0 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G]. \end{aligned} \tag{4.3}$$

These marginal probabilities are of the form (4.2).

The parameters of interest $(\boldsymbol{\beta}_{1:J}, c_{1:J}, G)$ become identified if they can be written as functions of $\boldsymbol{q}$. The identification analysis developed in the context of the random-effects 1PL-G actually provides insight for the identification of $(\boldsymbol{\beta}_{1:J}, c_{1:J}, G)$. As a matter of fact, in the parametric case the identification analysis was based on three steps, each of them involving specific features:

1. STEPS 1 and 2 essentially depend on the parameter $\sigma$ indexing the distribution generating the person specific abilities. More specifically, the item parameters $\boldsymbol{\beta}_{1:J}$ are written as a function of $\boldsymbol{c}_{1:J}$ and $\sigma$, whereas $\sigma$ is written as a function of $c_1$ and $c_2$; see Sections 3.3.1 and 3.3.2.
2. STEP 3 first establishes that the item parameters $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J})$ can be written as a function of $\beta_1$ and $c_1$. Second, using STEPS 1 and 2, it is concluded that $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma)$ is a function of $c_1$; see Section 3.3.3.

Thus, STEPS 1 and 2 critically depend on the parametric hypothesis which is assumed on the distribution generating the random effects $\theta_i$'s, whereas a part of STEP 3 does not depend on it. Moreover, in the random-effects 1PL-G model, the guessing parameters can be written as a function of the marginal probabilities and, therefore, do not depend on any specific distribution $G$; see Theorem 2, equality (3.16). These considerations suggest that STEP 3 can be used in the identification analysis of $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$.

*4.2.1. Identification of the Item Parameters* More precisely, let $J \geq 3$, $j \neq 1$ and $k \neq j$ (with $k, j \leq J$). Similarly to Section 3.3.3, we define the identified parameters $p_0^J$, $p_j^J$, and $p_{jk}^J$ as a function of the marginal probabilities given $(\boldsymbol{\beta}_{1:J}, \boldsymbol{\delta}_{1:J}, G)$. Thus, for instance, $p_0^J$ is defined as

$$p_0^J \doteq P\left[\bigcap_{1\leq j\leq J} \{Y_{ij} = 0\} \,\middle|\, \boldsymbol{\beta}_{1:J}, \boldsymbol{\delta}_{1:J}, G\right] = \prod_{1\leq j\leq J} \delta_j \times I_0^J(\boldsymbol{\beta}_{1:J}, G),$$

where $I_0^J(\boldsymbol{\beta}_{1:J}, G) = \int_{\mathbb{R}} \prod_{1\leq j\leq J} \frac{1}{1+e^{\theta-\beta_j}} G(d\theta)$; similarly, for $p_j^J$ and $p_{jk}^J$. Taking into account that, for each $j = 1, \ldots, J$, $\delta_j > 0$ and $\beta_j \in \mathbb{R}$, it follows, by the same arguments developed in Section 3.3.3, that

$$\delta_j = \frac{u_j \delta_1}{v_j \delta_1 + 1}, \qquad \beta_j = \beta_1 + \ln(v_j \delta_1 + 1), \quad j = 2, \ldots, J, \tag{4.4}$$

where $u_j$ and $v_j$ are defined as in Propositions 2 and 3 (with the corresponding change of notation). Since $u_j$'s and $v_j$'s are identified parameters, the following theorem can be established.

**Theorem 3.** *For the statistical model (4.2) induced by the semi-parametric 1PL-G model, the item parameters $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J})$ are identified by $\boldsymbol{Y}_1$ provided that*

1. *At least three items are available.*
2. $\beta_1 = 0$ *and* $c_1 = 0$.

*Moreover, under these identification restrictions, the item parameters can be expressed in terms of marginal probabilities; that is,*

$$\begin{aligned}
&\text{(i)} \quad \beta_j = \ln(v_j + 1) = \ln\left[\frac{p_j^J p_{1k}^J - p_1^J p_{jk}^J}{p_0^J p_{jk}^J - p_j^J p_k^J} + 1\right], \quad \textit{for } k \neq 1, j;\\
&\text{(ii)} \quad c_j = 1 - \frac{u_j}{v_j + 1} = 1 - \frac{p_0^J p_{1k}^J - p_1^J p_k^J}{p_{jk}^J (p_0^J - p_1^J) + p_j^J (p_{1k}^J - p_k^J)}, \quad \textit{for } k \neq 1, j.
\end{aligned} \tag{4.5}$$

This theorem deserves the following comments:

1. Under condition (4.4), the statistical model can be written as follows: for every $\mathcal{K} \subset \{1, \dots, J\}$,

$$\begin{aligned}
p_{\mathcal{K}} &\doteq P\left[\bigcap_{j \in \mathcal{K}} \{Y_{ij} = 0\} \,\middle|\, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G\right]\\
&= \prod_{j \in \mathcal{K}} u_j \times \int_{\mathbb{R}} \frac{G(d\theta)}{\prod_{j \in \mathcal{K}} \{v_j + \frac{1}{1-c_1} + \frac{e^{-\beta_1}}{1-c_1} e^{\theta}\}}\\
&= P\left[\bigcap_{j \in \mathcal{K}} \{Y_{ij} = 0\} \,\middle|\, \beta_1, c_1, G\right].
\end{aligned} \tag{4.6}$$

   Thus, the sampling process depends on the parameters of interest $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$ through $(\beta_1, c_1, G)$ only. Therefore, both the item parameters and the guessing parameters become identified if $\beta_1$ and $c_1$ are fixed at 0. Regarding the identification of $G$, see Section 4.2.2.
2. Condition (2) of Theorem 3 implies that the marginal probability that a person $i$ correctly answer the standard item 1 is given by

$$P[Y_{i1} = 1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G] = \int_{\mathbb{R}} \frac{\exp(\theta)}{1 + \exp(\theta)} G(d\theta).$$

   As in the random-effects case (see Section 3.3.4), this means that the measurement instrument must contain an item (labeled here with 1) such that each person answers it *without* guessing.
3. As in the random-effects 1PL-G model, the identification of the guessing parameters is obtained *without* depending on specific properties of the distribution $G$; see Equations (3.16) and (4.5.ii). Additionally, in the semi-parametric case, the identification of the difficulty parameters is also obtained independent of $G$; see Equation (4.5.i). However, in the random-effects 1PL-G model, the item parameters depend on the scale parameter of $G$, and consequently, on the dependency between two items; see Equations (3.11) and (3.12).
4. It should also be mentioned that equalities (4.5) provide an explicit statistical meaning for the item parameters in terms of the data generating process. This suggests to use them as estimators of $\beta_j$ and $c_j$ from the data through relative frequencies.

*4.2.2. Is the Distribution G Identified?* Let us now consider the possible identification of the distribution $G$ generating the random effects. Consider the $2^J - 1$ equations of the form (4.6); taking into account that $u_j$'s as well as $p_{\mathcal{K}}$ are identified, and using the identification restrictions established in Theorem 3, it follows that, for every subset $\mathcal{K} \subset \{1, \ldots, J\}$, except the empty set,

$$m_G(\mathcal{K}) \doteq \int_{\mathbb{R}} \frac{G(d\theta)}{\prod_{j\in\mathcal{K}}[v_j + 1 + e^{\theta}]} \tag{4.7}$$

is identified by the observations. Denote $\{m_G(\mathcal{K}) : \mathcal{K} \subset \{1, \ldots, J\} \setminus \emptyset\}$ as $\boldsymbol{m}_G$.

The question is to know whether the identification of these $2^J - 1$ functionals ensure the identification of $G$; that is, if two distributions $G_1$ and $G_2$ on $(\mathbb{R}, \mathcal{B})$ satisfy $\boldsymbol{m}_{G_1} = \boldsymbol{m}_{G_2}$, is it true that $G_1 = G_2$? We argue that those $2^J - 1$ equations are far from being enough to identify $G$. As a matter of fact, let us suppose that $G$ has a density function $g$ with respect to a $\sigma$-finite measure $\lambda$ on $(\mathbb{R}, \mathcal{B})$, that is, $g = dG/d\lambda$. Suppose furthermore that $g \in L^2(\mathbb{R}, \mathcal{B}, \lambda)$. Then (4.7) can be written as

$$m_g(\mathcal{K}) = \int_{\mathbb{R}} \frac{g(\theta)d\lambda(\theta)}{\prod_{j\in\mathcal{K}}[v_j + 1 + e^{\theta}]} \quad \forall \mathcal{K} \subset \{1, \ldots, J - 1\}, \text{ with } \mathcal{K} \neq \emptyset. \tag{4.8}$$

Since $v_j > -1$ for each $j$, it follows that

$$0 < \prod_{j\in\mathcal{K}} \frac{1}{v_j + 1 + e^{\theta}} \leq d \quad \forall \mathcal{K} \subset \{1, \ldots, J\}, \text{ with } \mathcal{K} \neq \emptyset$$

where $d$ is a positive real constant. Therefore,

$$f_{\mathcal{K}}(\theta) \doteq \prod_{j\in\mathcal{K}} \frac{1}{v_j + 1 + e^{\theta}} \in L^2(\mathbb{R}, \mathcal{B}, \lambda)$$

for each $\mathcal{K} \subset \{1, \ldots, J\}$, with $\mathcal{K} \neq \emptyset$.

Define the functional $T : L^2(\mathbb{R}, \mathcal{B}, \lambda) \longrightarrow \mathbb{R}^{2^J-1}$ as $Tg = \boldsymbol{m}_g$, where

$$\boldsymbol{m}_g = \big(m_g(\{1\}), m_g(\{2\}), \ldots, m_g(\{1, \ldots, J\})\big)$$

and $m_g(\mathcal{K})$ is, for each set $\mathcal{K}$, defined by (4.8). Thus, the identification of $g$ follows if

$$T(g_1 - g_2) = 0 \quad \Longrightarrow \quad g_1 = g_2.$$

If this is the case, then $g_1 - g_2 \in L^2(\mathbb{R}, \mathcal{B}, \lambda)$ is orthogonal to each $f_{\mathcal{K}}$, with the inner product

$$(f, h) = \int_{\mathbb{R}} f(\theta)h(\theta)\, d\lambda(\theta), \quad \text{for } f, h \in L^2(\mathbb{R}, \mathcal{B}, \lambda).$$

It follows that $g_1 - g_2$ is orthogonal to $\mathcal{N}$, the linear space generated by $\{f_{\mathcal{K}} : \mathcal{K} \subset \{1, \ldots, J\}, \mathcal{K} \neq \emptyset\}$. Since $\mathcal{N}$ is of finite dimension, it is therefore closed in $L^2(\mathbb{R}, \mathcal{B}, \lambda)$. Taking into account that $g_1 - g_2 = 0$ if and only if $(g_1 - g_2, f) = 0$ for all $f \in L^2(\mathbb{R}, \mathcal{B}, \lambda)$ (Halmos, 1951, Theorem 1, Section 4), if $g_1 - g_2 = 0$ then $\mathcal{N} = L^2(\mathbb{R}, \mathcal{B}, \lambda)$, which is impossible because $L^2(\mathbb{R}, \mathcal{B}, \lambda)$ is an infinite dimensional linear space.

Summarizing, we obtain the following theorem.

**Theorem 4.** *For the statistical model* (4.2) *induced by the semi-parametric* 1*PL-G model*, *assume that there are at least three items and that* $c_j > 0$ *for each* $j = 2, \ldots, J$. *Then the* $(2^J - 1)$-*dimensional vector* $\boldsymbol{m}_G$ *is identified by the observations*, *provided that* $\beta_1$ *and* $c_1$ *are fixed*, *but the distribution* $G$ *generating the individual abilities is not identified.*

It is relevant to inquire whether $\beta_1$ and/or $c_1$ can be identified if the mean of $\theta$ is fixed at 0. Taking into account Theorems 3 and 4, it can be stated that the only way to identify $\beta_1$ and/or $c_1$ through restrictions on $G$ is by restricting functionals of the type (4.8). More specifically, we known that the difficulty parameters and the guessing parameters are functions of both $(\beta_1, c_1)$ and the identified parameters $\{(u_j, v_j) : j = 2, \dots, J\}$. On the other hand, the only information of $G$ that is identified corresponds to the identified functionals

$$\widetilde{m}_G(\mathcal{K}) = \int_{\mathbb{R}} \frac{G(d\theta)}{\prod_{j\in\mathcal{K}} [v_j + \frac{1}{1-c_1} + \frac{e^{\theta-\beta_1}}{1-c_1}]} \quad \forall \mathcal{K} \subset \{1, \dots, J\}, \text{ with } \mathcal{K} \neq \emptyset. \tag{4.9}$$

Therefore, $\beta_1$ and $c_1$ become identified if at least two of these functionals are fixed at some specific values. For instance, we could choose the functionals $\widetilde{m}_G(\{2\})$ and $\widetilde{m}_G(\{3\})$ and to study if it is enough to solve them for $\beta_1$ and $c_1$. However, it is not clear which specific characteristics of $G$ are being fixed and, therefore, it is hard to propose this way as a solution to the identification problem. In any case, it is clear that the possibility of identifying $\beta_1$ and/or $c_1$ by fixing the mean of $\theta$ is far from being feasible since such a mean can not be written as a function of $\widetilde{m}_G(\mathcal{K})$ for some $\mathcal{K}$. The situation is different when $G$ is identified by the observations; see Theorem 5.

### *4.3. Identification Analysis Under an Infinite Quantity of Items*

The identification arguments previously developed, either in the parametric case or in the semi-parametric case, have a common feature, namely to write the parameters of interest as functions of identified parameters. Now the problem consists of writing the distribution $G$ as a function of identified parameters when an infinite number of items is available. These types of relationships can be developed in a Bayesian framework because in such a framework, the concept of identification reduces to a condition of measurability with respect to functionals of the sampling process.

*4.3.1. Bayesian Identification* In order to define Bayesian identification, it is necessary first to define the concept of a sufficient parameter.

**Definition 1.** Consider the Bayesian model defined by the joint probability distribution on $(\boldsymbol{Y}, \boldsymbol{\vartheta})$. A function $\boldsymbol{\psi} \doteq g(\boldsymbol{\vartheta})$ of the parameter $\boldsymbol{\vartheta}$ is a sufficient parameter for $\boldsymbol{Y}$ if the conditional distribution of the sample $\boldsymbol{Y}$ given $\boldsymbol{\psi}$ and $\boldsymbol{\vartheta}$ is the same as the distribution of the sample $\boldsymbol{Y}$ given $\boldsymbol{\psi}$, that is, $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\vartheta} \mid \boldsymbol{\psi}$.

Definition 1 implies that the sampling distribution generating $\boldsymbol{Y}$ is completely determined by the sufficient parameter $\boldsymbol{\psi}$, being $\boldsymbol{\vartheta}$ redundant. In fact, by definition of conditional independence, $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\vartheta} \mid \boldsymbol{\psi}$ implies that $E[h(\boldsymbol{Y}) \mid \boldsymbol{\vartheta}] = E[h(\boldsymbol{Y}) \mid \boldsymbol{\psi}]$ for every measurable function $h$. Equivalently, by the symmetry of a conditional independence relation, it can also be concluded that $\boldsymbol{\psi}$ is a sufficient parameter if the conditional distribution of the redundant part of $\boldsymbol{\vartheta}$, given the sufficient parameter $\boldsymbol{\psi}$, is not updated by the sample, that is, $E[f(\boldsymbol{\vartheta}) \mid \boldsymbol{Y}, \boldsymbol{\psi}] = E[f(\boldsymbol{\vartheta}) \mid \boldsymbol{\psi}]$ for every measurable function $f$.

Because of the numerous sufficient parameters in a given problem, one might ask whether one sufficient parameter $\boldsymbol{\psi}$ still contains redundant information about the sampling process. Suppose that there exists a function of $\boldsymbol{\psi}$, say $\boldsymbol{\psi}_1$, which is also a sufficient parameter for $\boldsymbol{Y}$, that is, $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\vartheta} \mid \boldsymbol{\psi}_1$. It follows that $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\psi} \mid \boldsymbol{\psi}_1$ and, therefore, the sampling process is fully characterized by $\boldsymbol{\psi}_1$, being $\boldsymbol{\psi}$ redundant in the sense that $E[h(\boldsymbol{Y}) \mid \boldsymbol{\vartheta}] = E[h(\boldsymbol{Y}) \mid \boldsymbol{\psi}] = E[h(\boldsymbol{Y}) \mid \boldsymbol{\psi}_1]$, for every measurable function $h$, or that $E[f(\boldsymbol{\psi}) \mid \boldsymbol{Y}, \boldsymbol{\psi}_1] = E[f(\boldsymbol{\psi}) \mid \boldsymbol{\psi}_1]$, for every measurable function $f$. Clearly, $\boldsymbol{\psi}_1$ should be preferred over $\boldsymbol{\psi}$ because it contains less redundant information about the sampling process. In fact, a parameterization that achieves the most parameter reduction while retaining all the information about the sampling process should be considered preferable. The definition of such a parameter is formalized next. For the Bayesian model defined on $(\boldsymbol{Y}, \boldsymbol{\vartheta})$, $\boldsymbol{\vartheta}_{\min}$ is a *minimal sufficient parameter* for $\boldsymbol{Y}$ if the following conditions are satisfied: (i) $\boldsymbol{\vartheta}_{\min}$ is a measurable function of $\boldsymbol{\vartheta}$, (ii) $\boldsymbol{\vartheta}_{\min}$ is a sufficient parameter, and (iii) for any other sufficient parameter $\boldsymbol{\psi}$, $\boldsymbol{\vartheta}_{\min}$ is a measurable function of it. From this definition, it follows that $\boldsymbol{\vartheta}_{\min}$ does not contain redundant information about the sampling process because there does not exist a non-injective function of it, say $\boldsymbol{\varphi}$, such that $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\vartheta}_{\min} \mid \boldsymbol{\varphi}$. These considerations lead to the definition of Bayesian identification (Florens & Rolin, 1984).

**Definition 2.** Consider the Bayesian model defined by the joint probability distribution on $(\boldsymbol{Y}, \boldsymbol{\vartheta})$. The parameter $\boldsymbol{\vartheta}$ is said to be Bayesian identified or $b$-identified if it is a minimal sufficient parameter.

The definition of $b$-identification depends on the prior distribution through its null sets only (for details, see, Florens et al., 1990; San Martín et al., 2011, Section 3). An important consequence of this is that unidentified parameters remain unidentified even if proper and concentrated priors are considered. Unidentified parameters can become identified if and only if the prior null sets are changed, which is equivalent to introducing dogmatic constraints.

From an operational point of view, it can be shown that the minimal sufficient parameter is generated by the family of all the sampling expectations $E[h(\boldsymbol{Y}) \mid \boldsymbol{\vartheta}]$, where $h \in L^1(\Omega, \mathcal{Y}, \Pi)$, with $L^1(\Omega, \mathcal{Y}, \Pi)$ being the set of integrable functions defined on the sample space $\Omega$ and measurable with respect to the $\sigma$-field $\mathcal{Y}$ generated by $\boldsymbol{Y}$; for details and proof, see Florens et al. (1990, Chapter 4). Let $\sigma\{E[h(\boldsymbol{Y}) \mid \boldsymbol{\vartheta}] : h \in L^1(\Omega, \mathcal{Y}, \Pi)\}$ be the $\sigma$-field generated by the sampling expectations. By definition of conditional expectation, this $\sigma$-field is contained in the $\sigma$-field generated by $\boldsymbol{\vartheta}$. Therefore, to show that $\boldsymbol{\vartheta}$ is $b$-identified by $\boldsymbol{Y}$, it is needed to show that $\boldsymbol{\vartheta}$ is measurable w.r.t. $\sigma\{E[h(\boldsymbol{Y}) \mid \boldsymbol{\vartheta}] : h \in L^1(\Omega, \mathcal{Y}, \Pi)\}$, or equivalently by the Doob–Dynkin lemma (see Rao, 2005, Chapter 2.1, Proposition 3), that $\sigma\{\boldsymbol{\vartheta}\} \subset \sigma\{E[h(\boldsymbol{Y}) \mid \boldsymbol{\vartheta}] : h \in L^1(\Omega, \mathcal{Y}, \Pi)\}$.

*4.3.2. Embedding Previous Identification Results in a Bayesian Framework* The identification analysis developed in Sections 3 and 4.2 is still valid under a Bayesian approach, provided that the parameters are endowed of proper prior distributions absolutely continuous with respect to the Lebesgue measure. As a matter of fact, let us consider the results established in Section 4.2. The corresponding Bayesian model is defined on $(\boldsymbol{Y}_1, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$, where the conditional distribution of $\boldsymbol{Y}_1$ given $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$ is of the form (4.2), and the parameters are endowed with a prior structure (which is detailed next). According to Section 4.3.1, the identified parameter corresponding to this model is given by $\sigma\{E[h(\boldsymbol{Y}_1) \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G] : h \in L^1(\{0,1\}^J, \mathcal{Y}, \Pi)\}$.

The goal is to prove under which conditions the parameters of interest $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$ are measurable with respect to $\sigma\{E[h(\boldsymbol{Y}_1) \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G] : h \in L^1(\{0,1\}^J, \mathcal{Y}, \Pi)\}$. The standard identification argument underlying Theorems 3 and 4 was to write the parameters of interest as a function of identified parameters; these parameters correspond to functionals of the sampling process. Thus, in Theorem 3, the difficulty parameters are a function of the identified parameters $\{v_j : j = 2, \ldots, J\}$. By construction, these parameters are functions of marginal probabilities of the form (4.2). Therefore, $\boldsymbol{\beta}_{2:J}$ is measurable with respect to $\sigma\{E[h(\boldsymbol{Y}_1) \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G] : h \in L^1(\{0,1\}^J, \mathcal{Y}, \Pi)\}$. Similarly, for the guessing parameters $\boldsymbol{c}_{1:J}$ and the functionals $\boldsymbol{m}_G$.

The identification results obtained in Theorems 3 and 4 can consequently be summarized in a Bayesian framework as follows (we include $\beta_1$ and $c_1$ since they are fixed at 0):

$$\sigma\{\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \boldsymbol{m}_G\} \subset \sigma\left\{E\left[h(\boldsymbol{Y}_1) \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G\right] : h \in L^1\left(\{0,1\}^J, \mathcal{Y}, \Pi\right)\right\} \subset \sigma\{\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G\}.$$

The first inclusion means that $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \boldsymbol{m}_G)$ are $b$-identified by $\boldsymbol{Y}_1$. The last inclusion shows that the parameters of interest $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$ are not $b$-identified by $\boldsymbol{Y}_1$; if it were, then

$$\sigma\{\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G\} \subset \sigma\{\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \boldsymbol{m}_G\},$$

which is impossible since $G$ can not be obtained by measurable operations from $\boldsymbol{m}_G$; see Section 4.2.2.

*4.3.3. Identification of the Parameters of Interest* According to Theorem 4, the functionals $\widetilde{m}_G(\mathcal{K})$ defined by (4.9) are $b$-identified by one observation. Defining $X = \frac{1}{1-c_1} + \frac{\exp(\theta-\beta_1)}{1-c_1}$, $\widetilde{m}_G(\mathcal{K})$ can be rewritten as

$$\widetilde{m}_G(\mathcal{K}) = \int_{\delta_1^{-1}}^{\infty} \frac{G_{\beta_1,c_1}(dx)}{\prod_{j\in\mathcal{K}}[v_j + x]}, \quad \mathcal{K} \subset \{1, \dots, J\} \setminus \emptyset,$$

where

$$G_{\beta_1,c_1}(x) = G\big[\beta_1 + \ln\big\{(1-c_1)x - 1\big\}\big]. \tag{4.10}$$

Note that the support of the random variable $X$ is $((1-c_1)^{-1}, \infty)$. Consider the following four conditions:

**H1.** For all $m, n \in \mathbb{N}$, $\boldsymbol{\theta}_{1:n}$, $\boldsymbol{\beta}_{1:m}$ and $\boldsymbol{c}_{1:m}$ are mutually independent conditionally on $(G, H, K)$.
**H2.** The item parameters $\boldsymbol{\beta}_{1:\infty}$ are generated by an *iid* process, where $H$ is the common probability distribution.
**H3.** The item parameters $\boldsymbol{c}_{1:\infty}$ are generated by an *iid* process, where $K$ is the common probability distribution.
**H4.** $G$, $H$ and $K$ are mutually independent.

The Bayesian model is now defined on $(\boldsymbol{Y}_1, \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G)$, where $\boldsymbol{Y}_1 \in \{0,1\}^{\mathbb{N}}$. Under these conditions, it can be proved that

$$\int_{\mathbb{R}^+} f(x) G_{\beta_1,c_1}(dx) \tag{4.11}$$

is $b$-identified by $\boldsymbol{Y}_1$, for every bounded continuous function $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$, where $G_{\beta_1,c_1}$ is defined by (4.10); that is, it is measurable with respect to $\sigma\{E[h(\boldsymbol{Y}_1) \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G] : h \in L^1(\{0,1\}^{\mathbb{N}}, \mathcal{Y}, \Pi)\}$. In particular, since

$$f_n(y) = \mathbb{1}_{(0,x]}(y) + \big[1 - n(y-x)\big]\mathbb{1}_{(x,x+\frac{1}{n})}(y) \downarrow \mathbb{1}_{(0,x]}(y) \quad \forall x \in \mathbb{R}^+,$$

as $n \to \infty$, the monotone convergence theorem implies that, for every $x \in \mathbb{R}^+$, $G_{\beta_1,c_1}((0,x])$, and so $G_{\beta_1,c_1}$, is identified by one observation.

It is interesting to remark how strong is the identification relationship (4.11) when it is compared with the identified functionals $\widetilde{m}_G(\mathcal{K})$: the first one is a condition valid for every function $f \in C_b(\mathbb{R}^+)$, where $C_b(\mathbb{R}^+)$ denotes the set of bounded continuous functions $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$, whereas the second condition is valid for the set

$$\left\{\prod_{j\in\mathcal{K}} [v_j + x]^{-1} : \mathcal{K} \subset \{1, \dots, J\} \setminus \emptyset, J \geq 3\right\} \subsetneq C_b\big((1-c_1)^{-1}, \infty\big) \subsetneq C_b\big(\mathbb{R}^+\big).$$

The following theorem, proved in Appendix C, establishes conditions under which the item parameters and the latent distribution $G$ are identified in the asymptotic Bayesian model $(\boldsymbol{Y}_1, \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G)$:

**Theorem 5.** *Consider an asymptotic Bayesian semi-parametric* 1*PL-G model obtained when the number of items* $J \to \infty$. *The item parameters* $(\boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty})$ *and the latent distribution generating the person specific abilities* $G$ *are b-identified by* $\boldsymbol{Y}_1$ *if the following two conditions hold*:

1. *The difficulty parameter* $\boldsymbol{\beta}_{1:\infty}$ *and the guessing parameters* $\boldsymbol{c}_{1:\infty}$ *satisfy conditions* **H1**–**H4**.
2. *At least one of the following identifying restrictions hold*:
   (a) $\beta_1 = 0$ *a.s. and* $c_1 = 0$ *a.s.*
   (b) $G$ *is a.s. a probability distribution on* $\mathbb{R}$ *such that its mean and variance are known constants*.
   (c) $G$ *is a.s. a probability distribution on* $\mathbb{R}$ *with two known* $q$*-quantiles*.

This theorem deserves the following comments:

1. Theorem 5 establishes the identification of the parameters of interest after fixing specific properties of the distribution $G$, for instance, its mean and variance. By so doing, $\beta_1$ and $c_1$ are identified. Contrary to Theorem 3, it is possible, when an infinite quantity of items is available, to identify $G_{\beta_1,c_1}$, and not only a "part" of it and, therefore, to choose which functionals can be restricted. For details, see Appendix C, STEP 4.
2. The proof of Theorem 5 is similar to the identification analysis of a semi-parametric Rasch model; see San Martín et al. (2011, Theorem 6). In fact, the structure of the proofs is the following:
   (a) First, the identified parameterization is made explicit. In both cases, the identified parametrization involves some parameters of interest: in the Rasch model case, it depends on one difficulty parameter; in the 1PL-G model case, it depends on both one difficulty parameter and one guessing parameter.
   (b) Second, the problem reduces to identifying a mixing distribution (in the Rasch model case, the mixing distribution is $G_{\beta_1}$, where $\beta_1$ corresponds to a difficulty parameter; in the 1PL-G case, the mixing distribution is $G_{\beta_1,c_1}$). Because of the binary support of the conditional distribution (in both cases, it is a Bernoulli distribution), the identification problem is solved asymptotically.

## 5. How Relevant is Identification in a Bayesian Approach?

The identification results established in Sections 2, 3, and 4, share a common identification restriction: the guessing parameter of the standard item should be fixed at 0. As it was discussed previously, this identification result imposes a design restriction on the multiple-choice test, namely to ensure that the test includes an item that no person will correctly answer by guessing. In practice, this means that not any kind of educational data can be analyzed with the 1PL-G model, but only those data which were generated by a measurement instrument satisfying the previous design.

Instead of limiting the applicability of the 1PL-G model to specific data sets, it could be argued that a Bayesian approach circumvents such a limitation since "unidentifiability causes no real difficulty in the Bayesian approach" (Lindley, 1971). However, when a model formalizes a certain phenomenon (in our case, the guessing behavior in multiple-choice tests), the identification problem is more than a simple technical assumption, but covers a more fundamental aspect, namely the adequacy of a theoretical statistical model for an observed process. Consequently, it is seems relevant to make clear the status of identification in a Bayesian framework.

### *5.1. Identified Parameter and Updating Process*

Lindley's statement is based on the fact that, under specific technical conditions (Mouchart, 1976; Florens et al., 1990, Chapter 1), it is possible to compute the posterior distribution of an unidentified parameter. By so doing, the unidentified parameter is revised by the observation and, therefore, has an empirical meaning. It is then concluded that "nonidentifiability does not assert that there is no Bayesian learning" (Gelgand & Sahu, 1999, p. 248). However, taking into account that a statistical model always includes an identified parameter, it can be shown that the posterior distribution of an unidentified parameter updates the identified parameter only, and consequently, does not provides any empirical information on the unidentified parameter.

As a matter of fact, let us consider, as in Section 4.3.1, a Bayesian model defined on $(\boldsymbol{Y}, \boldsymbol{\vartheta})$, where $\boldsymbol{Y}$ corresponds to the observations and $\boldsymbol{\vartheta}$ to the parameters. In the context of this Bayesian model, a sub-parameter $\boldsymbol{\zeta}$ is a measurable function of $\boldsymbol{\vartheta}$; we denote it as $\boldsymbol{\zeta} \subset \boldsymbol{\vartheta}$. Let $\boldsymbol{\vartheta}_{\min} \subset \boldsymbol{\vartheta}$ be the identified parameter and suppose that we are interested in revising an unidentified parameter $\boldsymbol{\zeta} \subset \boldsymbol{\vartheta}$. This leads to computing the posterior expectation $E[h(\boldsymbol{\zeta}) \mid \boldsymbol{Y}]$ for all measurable integrable function $h$.

What do we learn about $\boldsymbol{\zeta}$ from the data? By definition of $\boldsymbol{\vartheta}_{\min}$, we have that $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}_{\min}$. Since $\boldsymbol{\zeta} \subset \boldsymbol{\vartheta}$, it follows that

$$\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{\zeta} \mid \boldsymbol{\vartheta}_{\min}. \tag{5.1}$$

Using the standard properties of conditional expectation (Rao, 2005, Chapter 2), we obtain that, for all measurable integrable function $h$,

$$\begin{aligned} E\big[h(\boldsymbol{\zeta}) \mid \boldsymbol{Y}\big] &= E\big\{E\big[h(\boldsymbol{\zeta}) \mid \boldsymbol{Y}, \boldsymbol{\vartheta}_{\min}\big] \mid \boldsymbol{Y}\big\} \\ &= E\big\{E\big[h(\boldsymbol{\zeta}) \mid \boldsymbol{\vartheta}_{\min}\big] \mid \boldsymbol{Y}\big\} \qquad \text{by (5.1)} \\ &\doteq E\big[g(\boldsymbol{\vartheta}_{\min}) \mid \boldsymbol{Y}\big]. \end{aligned}$$

That is, what we learn about the unidentified parameter $\boldsymbol{\zeta}$ reduces to what we learn about a function of the identified parameter $\boldsymbol{\vartheta}_{\min}$. In other words, the data $\boldsymbol{Y}$ provides information on the identified parameter $\boldsymbol{\vartheta}_{\min}$ only, *not* on the unidentified parameter $\boldsymbol{\zeta}$: outside of the identified parameter $\boldsymbol{\vartheta}_{\min}$, there is nothing to learn.

The intuition behind this formal fact is the following: A parameter is Bayesian identified if and only if it is a measurable function of countably many sampling expectations of statistics. That is, the identified parameter captures all the information contained in the sampling process. This explains why an identification analysis needs to make explicit at the beginning the sampling distributions or likelihood.

*Example 1.* Let us illustrate the previous considerations with the following simple hierarchical model (taken from Poirier, 1998, Section 3.2): let $(Y \mid \psi, \lambda) \sim \mathcal{N}(\psi, \sigma_1^2)$ be the likelihood or sampling distribution, and let

$$\begin{pmatrix} \psi \\ \lambda \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_2^2 + 1 & 1 \\ 1 & 1 \end{pmatrix}\right)$$

be the prior specification. For simplicity, it is assumed that $\sigma_1^2$ and $\sigma_2^2$ are known constants. Since $\psi = E(Y \mid \psi, \lambda)$, it is concluded that $\psi$ is $b$-identified by $Y$, whereas $\lambda$ is unidentified. The posterior distribution of $\lambda$ can be computed and it is actually given by

$$(\lambda \mid Y) \sim \mathcal{N}\left(\frac{Y}{\sigma_1^2 + \sigma_2^2 + 1}, \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2 + 1}\right).$$

This distribution provides information about $\lambda$. However, it is a function of the posterior distribution of the identified parameter $\psi$. In fact, it is enough to note that

$$E[\lambda \mid Y] = \frac{1}{\sigma_2^2 + 1} E[\psi \mid Y], \qquad V(\lambda \mid Y) = \frac{1}{\sigma_2^2 + 1}\left[\frac{1}{\sigma_2^2 + 1} V(\psi \mid Y) + \sigma_2^2\right].$$

Therefore, the identified parameter fully characterizes the learning-by-observing process. Moreover, the statistical interpretation of the unidentified parameter rests on the statistical meaning of the identified one since the sampling process is fully characterized by the identified parameter.

Now, it is possible to marginalize with respect to $\psi$ and to declare that the sampling process is now described by the sampling distribution $(Y \mid \lambda) \sim \mathcal{N}(\lambda, \sigma_1^2 + \sigma_2^2)$, where the prior distribution is given by $\lambda \sim \mathcal{N}(0, 1)$. In this case, $\lambda$ is $b$-identified by $Y$ since $\lambda = E(Y \mid \lambda)$ and, therefore, fully characterizes the sampling process.

What is the sampling process we want to revise? This question should be solved by the modeler; once a sampling process is specified, the task, previous to any estimation procedure, is to know the adequacy of the theoretical statistical model for an observed process.

Taking into account these considerations, let us come back to the identification results established in the previous sections. Four different sampling processes have been made explicit and depending on substantive aspects, one of them should be chosen. For instance, if the person specific abilities are considered as parameters indexing the likelihood, along with the item parameters, then a fixed-effects 1PL-G model may be chosen. According to Theorem 1, under the restrictions $\beta_1 = 0$ and $c_1 = 0$, the identified parameters are $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J}, \boldsymbol{\theta}_{1:N})$. If the estimation procedure will be done using a Bayesian approach, the prior distributions defined on $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \boldsymbol{\theta}_{1:N})$ should be specified in such a way that $\beta_1 = 0$ a.s. and $c_1 = 0$ a.s. The identified parameters fully characterize the learning-by-observing process.

Suppose now that the interest focuses on the distribution generating the person specific abilities. In a realistic scenario, that is, when a finite number of items is available, the sampling process is parametrized by $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G)$; and, therefore, a semi-parametric 1PL-G model is considered. According to Theorems 3 and 4, if $\beta_1$ and $c_1$ are fixed at 0, then $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J}, \boldsymbol{m}_G)$ is $b$-identified by $\boldsymbol{Y}_1$. Thus, what we learn about $G$ from the data corresponds to the updating of $\boldsymbol{m}_G$ only. Therefore, it is not a matter of the non-parametric Bayesian procedure which is used to estimate $G$, the issue is that such an estimation always corresponds to $\boldsymbol{m}_G$, not to $G$.

### 5.2. *Identification, Consistency and the Gibbs Sampler*

The statistical model induced either by a random-effects 1PL-G model or by its semi-parametric version, corresponds to an *iid* process given the parameters of interest. In this context, the $b$-identified parameter is consistently estimated by the corresponding posterior expectation given $\boldsymbol{Y}_{1:N}$; for a proof, see Florens et al. (1990, Theorem 9.3.12). More precisely, for the random-effects 1PL-G model, $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma)$ is the $b$-identified parameter; then the following consistency result holds: for any measurable integrable function $f$,

$$\lim_{N\to\infty} E\big[f(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma) \mid \boldsymbol{Y}_{1:N}\big] = f(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma) \quad \text{a.s.}$$

In particular, the difficulty parameters $\boldsymbol{\beta}_{1:J}$, the guessing parameters $\boldsymbol{c}_{2:J}$ and the scale parameter $\sigma$ are consistently estimated by their posterior distributions. Similarly, for the semi-parametric 1PL-G model (with a finite quantity of items), $\boldsymbol{\beta}_{2:J}$, $\boldsymbol{c}_{2:J}$ and $\boldsymbol{m}_G$ are consistently estimated by their respective posterior expectations. Finally, for the semi-parametric 1PL-G model, the distribution $G$ is consistently estimated by its posterior expectation if both the number of items and the number of persons go to infinite. Let us mention that, for all prior distribution defined on the identified parameters, these results of consistency are necessary conditions for ensuring the consistency (both strong and weak) in a pure sampling theory framework; for general details, see Florens et al. (1990, Theorems 7.4.6 and 7.4.7).

In spite of the previous facts, the Bayesian literature has discussed the identification problem in relation to simulation-based techniques typically used for model fitting and inferences. It is commonly argued that nonidentifiability does not preclude Bayesian inference as long as a suitable informative prior is specified. Kass, Carlin, Gelman, and Neal (1998) pointed out that, provided the posterior is proper, there is no problem for MCMC methods, assuming that one has determined that the nonidentifiability "is not due to a bug". This type of consideration has been widely illustrated through the following example:

*Example 2.* Suppose that the data generating process is characterized by the sampling distribution

$$(Y \mid \psi, \lambda) \overset{\text{iid}}{\sim} \mathcal{N}\big(\psi + \lambda, \sigma_Y^2\big),$$

where $\sigma_Y^2$ is known. It is also assumed that $(\psi, \lambda)$ are the parameters of interest and, therefore, their prior distributions are specified as follows:

$$\big(\psi \mid \mu_\psi, \sigma_\psi^2\big) \sim \mathcal{N}\big(\mu_\psi, \sigma_\psi^2\big), \quad \big(\lambda \mid \mu_\lambda, \sigma_\lambda^2\big) \sim \mathcal{N}\big(\mu_\lambda, \sigma_\lambda^2\big), \quad \psi \perp\!\!\!\perp \lambda \mid \mu_\psi, \mu_\lambda, \sigma_\psi^2, \sigma_\lambda^2,$$

where $\mu_\psi, \mu_\lambda, \sigma_\psi^2, \sigma_\lambda^2$ are known constants. It is clear that $\psi + \lambda$ is the identified parameter, whereas $\psi$ and $\lambda$ are unidentified. Since these parameters are of interest, they are estimated through their posterior distribution. However, as in Example 1, these posterior distributions are functions of the posterior distribution of the identified parameter since

$$E[\psi \mid Y] = \eta_{\psi,\lambda} + \frac{\sigma_\psi^2}{\sigma_\lambda^2 + \sigma_\psi^2} E[\psi + \lambda \mid Y],$$

where $\eta_{\psi,\lambda} \doteq \frac{\sigma_\lambda^2}{\sigma_\lambda^2+\sigma_\psi^2}\mu_\psi - \frac{\sigma_\psi^2}{\sigma_\lambda^2+\sigma_\psi^2}\mu_\lambda$; and similarly for $E[\lambda \mid Y]$. In spite of this, it is said that if the prior distributions are not enough informative, the Gibbs sampler procedure, which is used to estimate the unidentified parameters, will show poor behavior; see Kass et al. (1998), Carlin and Louis (2000), Eberly and Carlin (2000), and Xie and Carlin (2006).

This type of statement is based on simulations such as the following. Using R (R Development Core Team, 2006), we simulated a sample of size 200 for a normal distribution with parameters $\psi = 5$ and $\lambda = 2$. We fit the model for three different prior specifications for $\psi$ and $\lambda$: Model 1: $\mathcal{N}(0, 1)$; Model 2: $\mathcal{N}(0, 5)$; Model 3: $\mathcal{N}(0, 10)$; Model 4: $\mathcal{N}(0, 100)$. For the Gibbs sampler procedure of all models, we generate 100,000 samples, discarding the initial 50,000 as the burn-in period and using a lag of 50 iterations to avoid correlation, for three different Markov chains, using different starting values for each chain. We report the posterior means, standard deviations and 95 % probability intervals in Table 1. In Figure 1, we show the running means of the unidentified parameter $\psi$, whereas in Figure 2 we show the Markov chain traces for $\psi$.

From Table 1 and Figures 1 and 2, it can be concluded that the Markov chain for the parameter $\psi$ converges only for Model 1, whereas for the other three models the convergence is getting worse as the variances $\sigma_\psi^2$ and $\sigma_\lambda^2$ increase. However, the Markov chain for the identified parameter $\psi + \lambda$ converges for the four models. This type of behavior leads to stating that if the prior distribution of the unidentified parameter is informative (Model 1), the Gibbs sampler procedure shows good behavior; and, therefore, an estimation of the unidentified parameter is obtained.

TABLE 1.
Bayesian estimates summaries for the simulated normal dataset.

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | $\psi$ | $\psi+\lambda$ | $\psi$ | $\psi+\lambda$ | $\psi$ | $\psi+\lambda$ | $\psi$ | $\psi+\lambda$ |
| Mean | 3.3 | 7.1 | 3.0 | 7.1 | 1.5 | 7.1 | 44.3 | 7.1 |
| SD | 0.7 | 0.1 | 2.3 | 0.1 | 9.3 | 0.1 | 34.7 | 0.1 |
| $P_{2.5}$ | 1.9 | 6.9 | −1.5 | 6.9 | −16.3 | 6.9 | −1.7 | 6.9 |
| $P_{97.5}$ | 4.7 | 7.2 | 7.2 | 7.2 | 18.4 | 7.2 | 124.6 | 7.2 |

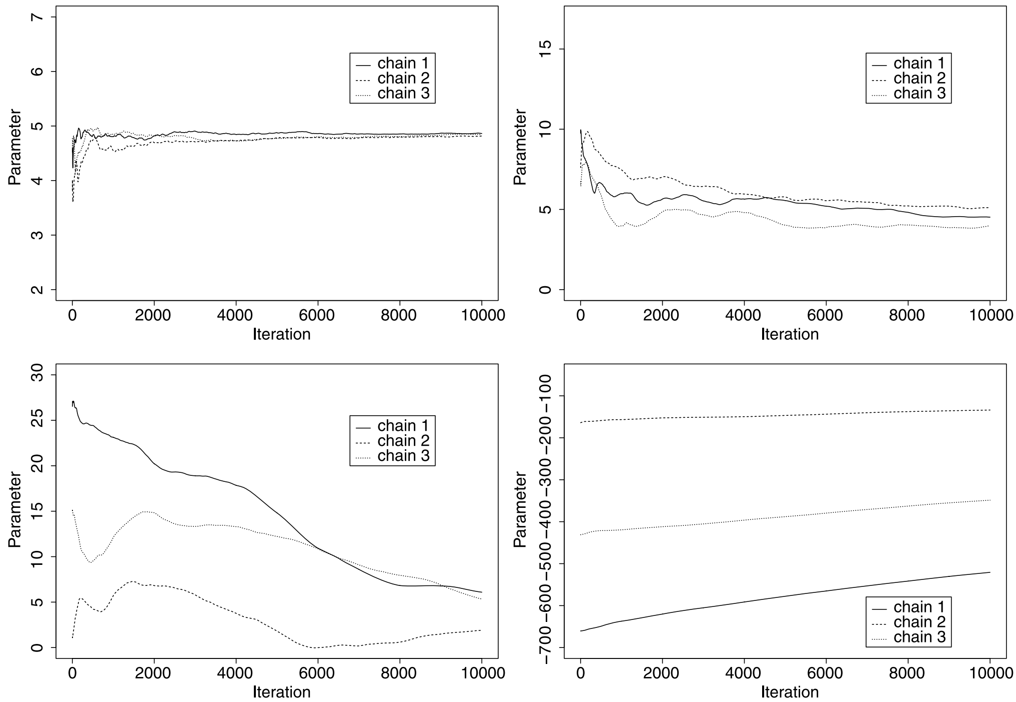


FIGURE 1.
Running means of parameter $\psi$: Model 1 (*upper left corner*), Model 2 (*upper right corner*), Model 3 (*lower left corner*), and Model 4 (*lower right corner*).

This type of conclusions should be complemented by standard convergence diagnostics. In applications, three convergence diagnostics are used: the Gelman and Rubin statistic (Gelman & Rubin, 1992); the Geweke test (Geweke, 1992); and the Heidelberg and Welch stationary test (Heidelberg & Welch, 1992). Tables 2 and 3 show the results of the convergence diagnostics for the Markov chains of the parameter $\psi$ and $\psi+\lambda$, respectively. It can be concluded that for Model 1, the Markov chain for the unidentified parameter $\psi$ converges according to these three criteria. However, it is important to stress that as the variances $\sigma_\psi^2$ and $\sigma_\lambda^2$ increase, the convergence of the Markov chain for the parameter $\psi$ is poor, whereas the Markov chain of the parameter $\psi+\lambda$ always has good behavior.

The conclusion seems to be the following: if prior distributions of unidentified parameters are concentrated around their true value, the Gibbs sampler procedure has good behavior. However, if those prior distributions become non-informative or diffuse (which happens when the variances $\sigma_\psi^2$ and $\sigma_\lambda^2$ are large), the Gibbs sampler procedure becomes worse.

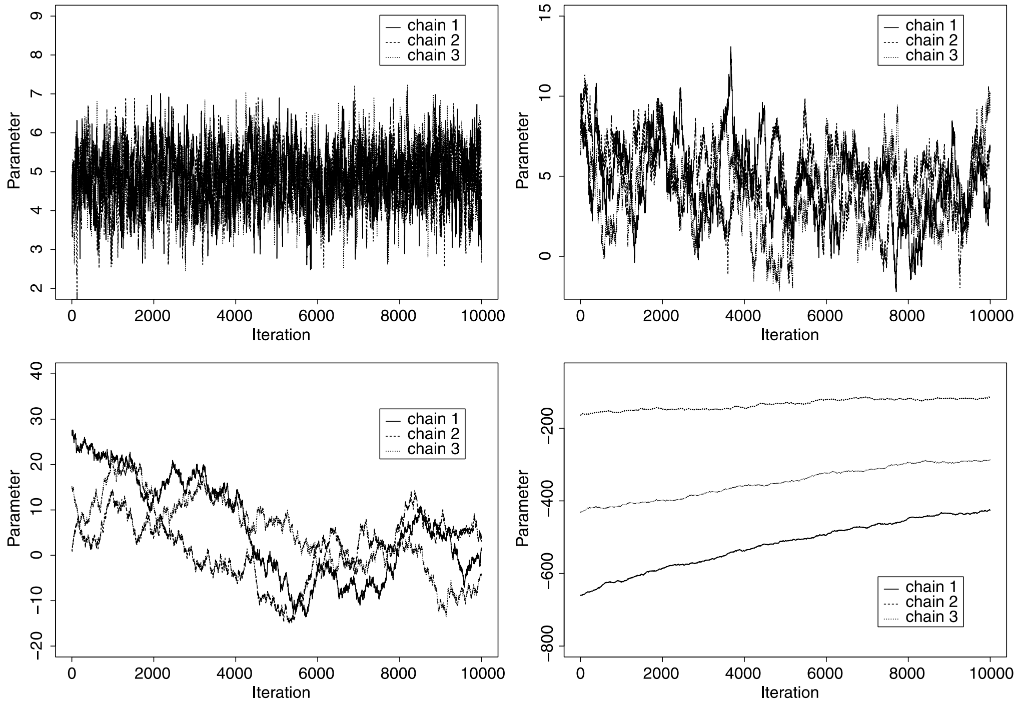


FIGURE 2.
Markov chain of parameter $\psi$: Model 1 (*upper left corner*), Model 2 (*upper right corner*), Model 3 (*lower left corner*), and Model 4 (*lower right corner*).

TABLE 2.
Convergence diagnostics for parameter $\psi$.

| Convergence diagnostic | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Gelman–Rubin (value of statistic) | 1.00 | 1.00 | 2.9 | 3.7 |
| Geweke (converging chains/total chains) | 3/3 | 0/3 | 0/3 | 0/3 |
| Heidelberg–Welch (converging chains/total chains) | 3/3 | 0/3 | 0/3 | 0/3 |

TABLE 3.
Convergence diagnostics for parameter $\psi + \lambda$.

| Convergence diagnostic | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Gelman–Rubin (value of statistic) | 1.00 | 1.00 | 1.00 | 1.00 |
| Geweke (converging chains/total chains) | 3/3 | 3/3 | 3/3 | 3/3 |
| Heidelberg–Welch (converging chains/total chains) | 3/3 | 3/3 | 3/3 | 3/3 |

The erratic behavior of the Markov chain is not related with the unidentifiability of the parameter $\psi$, but with its ergodicity. More precisely, let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X$ and $Y$ be two real random variables defined on $(\Omega, \mathcal{F})$. In order to define a two-component Gibbs sampler, we define a joint probability $\Pi$ as $\Pi(dx, dy) \doteq P[X \in dx, Y \in dy]$. It is assumed that $\Pi$ can be decomposed as follows:

$$\Pi(dx, dy) = \mu(dx)\nu^x(dy) = \nu(dy)\mu^y(dx),$$

where $\nu^{\bullet}(\bullet) : \mathbb{R} \times \mathcal{B} \longrightarrow [0,1]$ and $\mu^{\bullet}(\bullet) : \mathbb{R} \times \mathcal{B} \longrightarrow [0,1]$ are regular conditional probabilities, whereas $\mu(dx) = P[X \in dx]$ and $\nu(dy) = P[Y \in dy]$ are marginal probabilities. As is well known, a two-component Gibbs sampling algorithm is defined as follows:

1. A real random variable $X_0$ is generated from $\mu_0$: $X_0 \sim \mu_0$.
2. Given $X_0$, a real random variable $Y_0$ is generated as follows: $(Y_0 \mid X_0) \sim \nu^{X_0}$.
3. Given $(\boldsymbol{X}_{0:n}, \boldsymbol{Y}_{0:n})$, a real random variable $X_{n+1}$ is generated as follows: $(X_n \mid \boldsymbol{X}_{0:n}, \boldsymbol{Y}_{0:n}) \sim \mu^{Y_n}$ for all $n \in \mathbb{N}$; and given $(\boldsymbol{X}_{0:n+1}, \boldsymbol{Y}_{0:n})$, a real random variable $Y_{n+1}$ is generated as follows: $(Y_{n+1} \mid \boldsymbol{X}_{0:n+1}, \boldsymbol{Y}_{0:n}) \sim \nu^{X_{n+1}}$ for all $n \in \mathbb{N}$.

Here $\boldsymbol{X}_{0:n} = (X_0, X_1, \ldots, X_n)$ and similarly for $\boldsymbol{Y}_{0:n}$. The following two properties are known: let $Z_n = (X_n, Y_n)$ and $\boldsymbol{Z} \doteq \{Z_n : n \in \mathbb{N}\}$. Then

1. $\boldsymbol{Z}$ is a homogeneous Markov process $P^{\mu_0}$.
2. If $\mu_0 = \mu$, then $Z$ is a stationary Markov process $P^{\mu}$.

In this context, it can be shown that the invariant $\sigma$-field, denoted as $\mathcal{Z}_I$, can be characterized by the following intersection of $\sigma$-fields: $\mathcal{Z}_I = \overline{\mathcal{X}}_0 \cap \overline{\mathcal{Y}}_0$, where $\overline{\mathcal{X}}_0$ corresponds to the $\sigma$-field generated by $X_0$ and completed by measurable sets of probability 0 or 1; similarly, for $\overline{\mathcal{Y}}_0$. Therefore, $\boldsymbol{Z}$ is an ergodic stationary Markov process if and only if

$$\overline{\mathcal{X}}_0 \cap \overline{\mathcal{Y}}_0 = \{A \in \mathcal{F} : P(A)^2 = P(A)\}. \tag{5.2}$$

Note that the $\sigma$-field at the right-hand side of (5.2) corresponds to the completed trivial $\sigma$-field. For details and proofs, see Florens et al. (1990, Theorem 9.3.24) and Berti, Pratelli, and Riggo (2008, 2010). This characterization can be extended to a $k$-component Gibbs sampler, but this is outside of the scope of this paper.

Let us comment on the equality (5.2). When the equality (5.2) holds, it is said that $\mathcal{X}_0$ and $\mathcal{Y}_0$ *are measurably separated* (Florens et al., 1990, Chapter 5). Taking into account that the $\sigma$-field generated by a random variable corresponds to the sets of events that may be described in terms of that random variable (Florens & Mouchart, 1982; San Martín, Mouchart, & Rolin, 2005), the completed trivial $\sigma$-field at the right-hand of (5.2) corresponds to the almost-sure trivial information. Therefore, (5.2) can heuristically be interpreted as saying that $X_0$ and $Y_0$ *don't share common information*. This concept is related to Basu's first theorem and to non-common information in graphical models; for details, see San Martín et al. (2005).

In the context of Example 2, the ergodicity of the Markov chain induced by the two-component Gibbs sampler corresponds to the measurable separability of the first two states of the chain $\boldsymbol{\theta}_0 = (\psi_0, \lambda_0)$ and $\boldsymbol{\theta}_1 = (\psi_1, \lambda_1)$ conditionally on $Y$. Using San Martín et al. (2005, Theorem 4.1), this condition is equivalent to the following condition:

$$r\big[Var(\boldsymbol{\theta}_1 \mid Y)\big] = r\big[Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y)\big], \tag{5.3}$$

where $r(A)$ denotes the rank of matrix $A$. Assuming that $\sigma_Y = 1$, it can be verified that

$$Var(\boldsymbol{\theta}_1 \mid Y) = \frac{1}{(\sigma_\psi^2 + \sigma_\lambda^2 + 1)} \begin{pmatrix} \sigma_\psi^2(\sigma_\lambda^2 + 1) & -\sigma_\lambda^2\sigma_\psi^2 \\ -\sigma_\lambda^2\sigma_\psi^2 & \sigma_\lambda^2(\sigma_\psi^2 + 1) \end{pmatrix}$$

and

$$Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y) = \begin{pmatrix} \frac{\sigma_\psi^2}{\sigma_\psi^2+1} & \frac{-\sigma_\lambda^2\sigma_\psi^2}{(\sigma_\psi^2+1)(\sigma_\lambda^2+1)} \\ \frac{-\sigma_\lambda^2\sigma_\psi^2}{(\sigma_\psi^2+1)(\sigma_\lambda^2+1)} & \frac{\sigma_\lambda^2}{\sigma_\lambda^2+1}[1 + \frac{\sigma_\psi^2\sigma_\lambda^2}{(\sigma_\psi^2+1)(\sigma_\lambda^2+1)}] \end{pmatrix}.$$

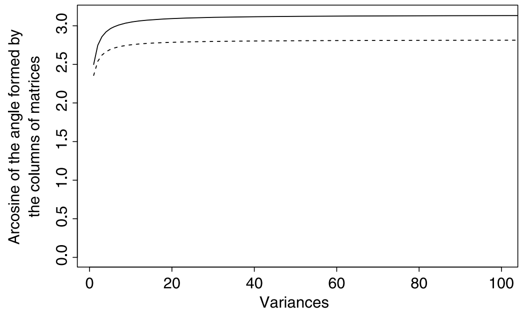


FIGURE 3.
Arccosine of the angle formed by the columns of matrix $Var(\boldsymbol{\theta}_1 \mid Y)$ (*continuous line*) and $Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y)$ (*dash line*).

It is straightforward to verify that these two matrices are of rank 2 and, therefore, (5.3) is verified. In other words, the Markov chain induced by the two-component Gibbs sampler is theoretically ergodic; and, consequently, no convergence problem exists. Nevertheless, this fact is actually contrary to the conclusions based on the convergence diagnostics. The problem arises when $\sigma_\psi^2$ and $\sigma_\lambda^2$ go to infinity. In this case, we have that

$$Var(\boldsymbol{\theta}_1 \mid Y) \longrightarrow \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y) \longrightarrow \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

This means that as the variances $\sigma_\psi^2$ and $\sigma_\lambda^2$ increase, the rank of matrix $Var(\boldsymbol{\theta}_1 \mid Y)$ *numerically* approaches 1, whereas the rank of $Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y)$ is always equal to 2. A graphical view of this empirical fact is provided in Figure 3, where we plot a common value for the variances $\sigma_\psi^2$ and $\sigma_\lambda^2$ against the arccosine of the angle formed by the columns of $Var(\boldsymbol{\theta}_1 \mid Y)$ and $Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y)$. It can be appreciated that as the variances grow, the angle formed by the columns of $Var(\boldsymbol{\theta}_1 \mid Y)$ becomes increasingly closer to 180° (or $\pi$ radians), which means that both columns are "almost" linearly dependent. However, as the variances grow, the angle formed by the columns of $Var(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_0, Y)$ becomes increasingly closer to 161.5° (or $\approx 9\pi/10$ radians), which means that both columns are always linearly independent. Therefore, the erratic behavior of the Gibbs sampler is due to the fact that the ergodicity of the Markov chain of Example 2 empirically seems to fail, although theoretically it is always ergodic. This suggests that is necessary to generate more samples in order to observe the convergence.

The measurable separability of the first two states of the Markov chain induced by the Gibbs sampler is a necessary and sufficient condition for the a.s. convergence of the chain. Therefore, the behavior of the Gibbs sampler does not depend on the (un)identification of parameters, but only on the information common to the first two states of the corresponding Markov chain. Identification ensures that the results obtained by the Gibbs sampler correspond to the estimation of the identified parameter. Moreover, if an unidentified parameter is estimated through a Gibbs sampler procedure, its results must be interpreted as an estimation of the identified parameters. Let us finish this section mentioning that in future works we plan to address the problem of characterizing condition (5.2) for semi-parametric IRT models.

## 6. Discussion

We have studied the identification problem of a particular case of the 3PL model, namely the 1PL-G model which assumes that the discrimination parameters are all equal to 1. The identification problem was studied under three different specifications. The first specification assumes that the individual abilities are unknown parameters. The second specification considers the abilities as mutually independent random variables with a common distribution known up to the scale parameter. In this context, the case was also considered where the distribution generating the individual abilities is known up to the scale parameter and the location parameter. The third specification corresponds to a semi-parametric 1PL-G model, where the distribution generating the individual abilities is unspecified and, consequently, considered as a parameter of interest.

### 6.1. *Summary of the Main Results*

For the first specification, the parameters of interest are the difficulty parameters, the guessing parameters and the individual abilities. These are identified provided a difficulty parameter and a guessing parameter are fixed at zero. For the second specification, the parameters of interest are the difficulty and guessing parameters, and the scale parameter. It was shown that these parameters are identified by one observation if a guessing parameter is fixed at zero. Also studied was the identification problem when the distribution generating the individual abilities is known up to both the scale and the location parameters. In this context, the parameters of interest are identified if a guessing parameter and a difficulty parameter are fixed at 0. As a sub-product of the previous arguments, the parameters of interest of a random-effects 2PL model are identified provided a discrimination parameter is fixed at 1.

For the third specification, the parameters of interest are the difficulty parameters, the guessing parameters and the distribution $G$ generating the individual abilities. When at least three items are available, the item parameters are identified provided a difficulty parameter and a guessing parameter are fixed at 0. However, under these identification restrictions, it was proved that the distribution $G$ is not identified. This lack of identification jeopardizes the empirical meaning of an estimate for $G$ under a finite number of items. In the unrealistic case when an infinite quantity of items is available, the distribution $G$ and the item parameters become identified if either a difficulty and guessing parameters are fixed at zero, or two characteristics of $G$ (the first two moments, or two quantiles) are fixed. For an overview of these results, see Table 4.

It should be remarked that the proofs of these identification results consisted in obtaining the corresponding identified parameterizations of the sampling process. Thereafter, identification restrictions were imposed in such a way that the parameters of interest become identified. This means that the identification restrictions are not only sufficient conditions, but also necessary. On the other hand, these results show that the identification of the fixed-effects 1PL-G model *does not* imply the identification of the random-effects 1PL-G model and, by extension, of the semi-parametric 1PL-G model.

### 6.2. *Practical Consequences of the Main Results*

Among the practical consequences of the previous identification results, we remark the following:

1. The identification results established in Sections 2, 3, and 4, share a common identification restriction (except Corollary 1, second statement): the guessing parameter of the standard item should be fixed at 0. This result imposes a design restriction on the multiple-choice test, namely to ensure that the test includes an item that no person will correctly answer by guessing. In practice, this means that not any kind of educational data can be analyzed with the 1PL-G model, but only those data which were generated by a multiple-choice test satisfying the previous design. The investigation of guessing behavior in multiple-choice tests requires, therefore, the comprehension of what it means to answer correctly an item without guessing. This is a theoretical challenge that should be satisfied if the 1PL-G model (as well as other IRT models with a guessing parameter) wants to be fitted.

TABLE 4.
Identification restrictions in different specifications of the 1PL-G model.

| Model specification | Parameters of interest | Identification restrictions | Number of items | Identified parameters | Theorem |
|---|---|---|---|---|---|
| Fixed-effects | $(\boldsymbol{\theta}_{1:N}, \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}) \in \mathbb{R}^N \times \mathbb{R}^J \times [0,1]^J$ | (i) $\beta_1 = 0$ and $c_1 = 0$;<br>(ii) $N \geq 2$;<br>(iii) $P[Y_{ij} = 1 \mid \theta_i, \beta_j, c_j] \neq P[Y_{i'j} = 1 \mid \theta_{i'}, \beta_j, c_j]$ for every $j$, where $i \neq i'$. | $1 \leq J < \infty$ | $(\boldsymbol{\theta}_{1:N}, \boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J})$ | Theorem 1 |
| Random-effects with $(\theta_i \mid \sigma) \sim G^\sigma$ | $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma) \in \mathbb{R}^J \times [0,1]^J \times \mathbb{R}^+$ | (i) $c_1 = 0$. | $3 \leq J < \infty$ | $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{2:J}, \sigma)$ | Theorem 2 |
| Random-effects with $(\theta_i \mid \sigma, \mu) \sim G^{\mu,\sigma}$ | $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \mu, \sigma) \in \mathbb{R}^J \times [0,1]^J \times \mathbb{R} \times \mathbb{R}^+$ | (i) $c_1 = 0$;<br>(ii) $\boldsymbol{a}'\boldsymbol{\beta}_{1:J} = 0$, with $\mathbb{1}'_J\boldsymbol{a} \neq 0$ and $\boldsymbol{a}$ known. | $3 \leq J < \infty$ | $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J}, \mu, \sigma)$ | Corollary 2 |
| Semi-parametric | $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G) \in \mathbb{R}^J \times [0,1]^J \times \mathcal{P}(\mathbb{R}, \mathcal{B})$ | (i) $\beta_1 = 0$ and $c_1 = 0$. | $3 \leq J < \infty$ | $(\boldsymbol{\beta}_{2:J}, \boldsymbol{c}_{2:J}, \boldsymbol{m}_G)$ | Theorems 3 and 4 |
| Semi-parametric | $(\boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, G) \in \mathbb{R}^J \times [0,1]^J \times \mathcal{P}(\mathbb{R}, \mathcal{B})$ | (i) **H1–H4**;<br>(ii) At least one of the following conditions hold:<br>(iii.1) $\beta_1 = 0$ and $c_1 = 0$.<br>(iii.2) The variance and the mean of $G$ are a.s. known.<br>(iii.3) Two $q$-quantiles of $G$ are a.s. known. | $J = \infty$ | $(\boldsymbol{\beta}_{2:\infty}, \boldsymbol{c}_{2:\infty}, G)$ under (iii.1); or $(\boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G)$ under (iii.2) or (iii.3), except what it is known on $G$. | Theorem 5 |

2. For the random-effects 1PL-G model, the identification analysis implies that the guessing parameters are bounded by the marginal probability of answering an item correctly; see inequality (3.15). Taking into account that the identification analysis is done under the assumption that the specific statistical model is the true model, in practice this condition could be useful for evaluating the adequacy of the 1PL-G model for the data under analysis. In fact, first, $c_j$ should be approximated by (3.16) for each item $k \neq 1, j$; secondly, if there exists an item (other than the standard item 1) $j$ such that $P[Y_{ij} = 1 \mid \boldsymbol{\beta}_{1:J}, \boldsymbol{c}_{1:J}, \sigma] < c_j$, for one item $k \neq 1, j$, then we have some empirical evidence to reject the use of this model. In any case, empirical investigations need to be performed to take practical advantages of these relationships.
3. Inequality (3.15) can also be used to evaluate if the practical rules of fixing an overall guessing parameter at $L^{-1}$, with $L$ being the number of response categories, is empirically adequate. In other words, this inequality is useful for interpreting the guessing behavior with respect to $L^{-1}$ in the sense that people guess considering all the alternatives or some of them. In the last case, it could be suggested that the guessing behavior is ability-based.
4. In applications, it is well known that the 3PL has estimation problems, specially for the guessing parameters. Some software prevent against this type of problems; see, for instance, Rizopoulos (2006). Taking into account that the 1PL-G model is a particular case of the 3PL model, our results explain that a source of such problems is due to the lack of parameter identification. Therefore, if specialized packages are used to fit the 1PL-G model, the identification restrictions established in this paper should be considered in order to ensure a coherent inference.
5. In the semi-parametric case, when a finite number of items is available (which is always the case), it was established that the distribution $G$ generating the person specific abilities is not identified. This lack of identification jeopardizes the empirical meaning of an estimate for $G$ under a finite number of items. This result is of practical relevance, especially considering the large amount of research trying to relax the parametric assumption of $G$ in the IRT literature.

Let us finish this paper by pointing out two open problems we consider of relevance. First, the identification of the random-effects 1PL-G model depends on the logistic function used in its specification, although part of the identification analysis was developed for general link functions $F$. Is it possible to extend the identification analysis for arbitrary link functions $F$? Second, the identification analysis exploited in this paper could be useful for investigating the identification of both the fixed-effects and the random-effects 3PL model. In this paper, not only was the identification of the random-effects 1PL-G model established, but also the identification of the random-effects 2PL model. Taking into account that these models are related to the 3PL, it seems reasonable to expect that the identification of the 3PL model is based on them. However, this is still an open problem.

## Acknowledgements

The work developed in this paper was presented in a Symposium on Identification Problems in Psychometrics at the *International Meeting of the Psychometric Society IMPS 2009*. The meeting was held in Cambridge (UK), in July 2009. The first author gratefully acknowledges the partial financial support from the ANILLO Project SOC1107 from the Chilean Gouvernment. The third author acknowledges the partial financial support from the Grant FONDECYT 11100076 from Chilean Government. The authors gratefully acknowledge several discussions with Claudio Fernández (Faculty of Mathematics, Pontificia Universidad Católica de Chile) and

Paul De Boeck (University of Amsterdam). This paper benefited from the helpful suggestions of three anonymous referees and the associate editor. In particular, one of the questions proposed by a referee led us to correct an error in a conclusion of Theorem 2.

## Appendix A. Identifiability of the Scale Parameter $\sigma$ by $\omega_{12}, \delta_1, \delta_2, \omega_1$, and $\omega_2$

To prove that the function $\omega_{12}$ given by (3.11) 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 using the Implicit Function Theorem (Spivak, 1965), but also assuming regularity conditions that allow performing derivatives under the integral sign. We accordingly assume that the cumulative distribution function $F$ has a continuous density function $f$ strictly positive on $\mathbb{R}$. Furthermore, to prove that $\omega_{12}$ 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 (3.7) with respect to $\sigma$ and to $\beta$. Consequently, it is assumed that, $\forall \sigma \in \mathbb{R}_0^+$ and $\forall \beta \in \mathbb{R}$, there exist $\epsilon > 0$ and $\eta > 0$, such that

$$\int_{\mathbb{R}} \sup_{|\sigma'-\sigma|\leq\epsilon} \sup_{|\beta'-\beta|\leq\eta} f\big(\sigma' x - \beta'\big) G(dx) < \infty,$$

$$\int_{\mathbb{R}} |x| \sup_{|\sigma'-\sigma|\leq\epsilon} \sup_{|\beta'-\beta|\leq\eta} f\big(\varphi' x - \beta'\big) G(dx) < \infty.$$

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

$$\begin{aligned} &\text{(i)} \quad D_2 p(\sigma,\beta) \doteq \frac{\partial}{\partial\beta} p(\sigma,\beta) = \int_{\mathbb{R}} f(\sigma x - \beta) G(dx) \\ &\text{(ii)} \quad D_1 p(\sigma,\beta) \doteq \frac{\partial}{\partial\sigma} p(\sigma,\beta) = -\int_{\mathbb{R}} x f(\sigma x - \beta) G(dx). \end{aligned} \tag{A.1}$$

Thus, $\overline{p}(\sigma, \alpha)$ as defined by (3.9) is also continuously differentiable on $\mathbb{R}_0^+ \times (0, 1)$; and from (3.10), we obtain that

$$\begin{aligned} \text{(i)} \quad 1 &= \frac{\partial}{\partial\beta} \overline{p}\big[\sigma, p(\sigma,\beta)\big] \\ &= D_2\overline{p}\big[\sigma, p(\sigma,\beta)\big] \times D_2 p(\sigma,\beta) \\ \text{(ii)} \quad 0 &= \frac{\partial}{\partial\sigma} \overline{p}\big[\sigma, p(\sigma,\beta)\big] \\ &= D_1\overline{p}\big[\sigma, p(\sigma,\beta)\big] + D_2\overline{p}\big[\sigma, p(\sigma,\beta)\big] \times D_1 p(\sigma,\beta), \end{aligned} \tag{A.2}$$

where

$$D_1\overline{p}(\sigma,\omega) \doteq \frac{\partial}{\partial\sigma}\overline{p}(\sigma,\omega), \qquad D_2\overline{p}(\sigma,\omega) \doteq \frac{\partial}{\partial\omega}\overline{p}(\sigma,\omega).$$

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

$$\begin{aligned} \text{(i)} \quad D_2\overline{p}(\sigma,\omega) &= \frac{1}{D_2 p[\sigma, \overline{p}(\sigma,\omega)]} = \frac{1}{\int_{\mathbb{R}} f[\sigma x - \overline{p}(\sigma,\omega)] G(dx)} \\ \text{(ii)} \quad D_1\overline{p}(\sigma,\omega) &= -\frac{D_1 p[\sigma, \overline{p}(\sigma,\omega)]}{D_2 p[\sigma, \overline{p}(\sigma,\omega)]} = \frac{\int_{\mathbb{R}} x f[\sigma x - \overline{p}(\sigma,\omega)] G(dx)}{\int_{\mathbb{R}} f[\sigma x - \overline{p}(\sigma,\omega)] G(dx)} \\ &\doteq E_{\sigma,\omega}(X), \end{aligned} \tag{A.3}$$

where

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

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 $\omega_{12}$ is continuously differentiable under the integral sign in $\sigma$, $\beta_1$ and $\beta_2$; therefore, the function $\varphi(\sigma,\delta_1,\delta_2,\omega_1,\omega_2)$ is continuously differentiable under the integral sign with respect to $\sigma$. It remains to show that the derivative w.r.t. $\sigma$ is strictly positive. Let us consider the sign of one of the two terms of the derivative of $\varphi(\sigma,\delta_1,\delta_2,\omega_1,\omega_2)$. Using (A.3.ii), we obtain that

$$\frac{\partial}{\partial\sigma}\bigl\{1 - F\bigl[\sigma x - \overline{p}(\sigma,\omega_1/\delta_1)\bigr]\bigr\} = -f\bigl[\sigma x - \overline{p}(\sigma,\omega_1/\delta_1)\bigr]\bigl(x - E_{\sigma,\omega_1/\delta_1}(X)\bigr).$$

Therefore, such a second term can be written as

$$\begin{aligned}
&\delta_1\delta_2\int_{\mathbb{R}} -f\left[\sigma\theta - \overline{p}\left(\sigma,\frac{\omega_1}{\delta_1}\right)\right]\bigl\{\theta - E_{\sigma,\omega_1/\delta_1}(\theta)\bigr\} \times \left\{1 - F\left[\sigma\theta - \overline{p}\left(\sigma,\frac{\omega_2}{\delta_2}\right)\right]\right\}G(d\theta)\\
&\quad = \delta_1\delta_2\int_{\mathbb{R}} f\left[\sigma\theta - \overline{p}\left(\sigma,\frac{\omega_1}{\delta_1}\right)\right]G(d\theta) \times C_{\sigma,\omega_1/\delta_1}\left\{\theta, F\left[\sigma\theta - \overline{p}\left(\sigma,\frac{\omega_2}{\delta_2}\right)\right]\right\}.
\end{aligned}$$

Now, since $F[\sigma\theta - \overline{p}(\sigma,\omega_2/\delta_2)]$ is a strictly increasing function of $\theta$, the covariance between $\theta$ and $F[\sigma\theta - \overline{p}(\sigma,\omega_2/\delta_2)]$ (with respect to $G_{\sigma,\omega_1/\delta_1}$) is strictly positive (if $\theta$ is not degenerate). Furthermore,

$$\int_{\mathbb{R}} f\bigl[\sigma x - \overline{p}(\sigma,\omega_1/\delta_1)\bigr]G(dx)$$

is clearly strictly positive. The two terms of the derivative of $\varphi(\sigma,\delta_1,\delta_2,\omega_1,\omega_2)$ are, therefore, strictly positive. □

## Appendix B. Identification of the Random-Effects 2PL Model

Random-effects 2PL-type models are specified under the same hypotheses of the random-effects 1PL-G model (see Section 3.1), but the conditional distribution of $Y_{ij}$ given the person specific ability $\theta_i$ is given by

$$P[Y_{ij} = 1 \mid \theta_i, \alpha_j, \beta_j] = F(\alpha_j\theta_i - \beta_j), \tag{B.1}$$

where $F$ is a strictly increasing distribution function with a continuous density function $f$ strictly positive on $\mathbb{R}$. Let us suppose that the person specific abilities are distributed according to a known distribution $G$.

### *B.1. Identification of the Difficulty Parameters*

Let

$$\gamma_j \doteq \int_{\mathbb{R}} F(\alpha_j\theta - \beta_j)G(d\theta) \doteq p(\alpha_j,\beta_j),$$

which is a continuous function strictly decreasing in $\beta_j$. Define

$$\overline{p}(\alpha,\gamma) = \inf\bigl\{\beta : p(\alpha,\beta) < \gamma\bigr\}.$$

Since $\overline{p}[\alpha, p(,\alpha,\beta)] = \beta$, it follows that $\beta_j = \overline{p}[\alpha_j,\gamma_j]$ for each $j = 1,\ldots,J$. Thus, the item parameter $\beta_j$ becomes identified once the discrimination parameter $\alpha_j$ becomes identified.

*B.2. Monotonicity of* $P[Y_{ij}=1, Y_{ik}=1 \mid \boldsymbol{\alpha}_{1:J}, \boldsymbol{\beta}_{1:J}]$

In order to identify the discrimination parameters, we need to study the monotonicity of $P[Y_{ij}=1, Y_{ik}=1 \mid \boldsymbol{\alpha}_{1:J}, \boldsymbol{\beta}_{1:J}]$ as a function of the discrimination parameters. Using the equality $\overline{p}[\alpha, p(\alpha,\beta)] = \beta$, it follows that

$$\frac{\partial}{\partial \zeta}\overline{p}\big[\alpha, p(\alpha,\beta)\big] = -\frac{1}{\int_{\mathbb{R}} f(\alpha\theta-\beta)G(d\theta)};$$
$$\frac{\partial}{\partial \alpha}\overline{p}\big[\alpha, p(\alpha,\beta)\big] = \frac{\int_{\mathbb{R}} \theta f(\alpha\theta-\beta)G(d\theta)}{\int_{\mathbb{R}} f(\alpha\theta-\beta)G(d\theta)} \doteq E_{\alpha,\beta}[X],$$

where

$$P_{\alpha,\beta}[X \in d\theta] \doteq G_{\alpha,\beta}(d\theta) = \frac{f(\alpha\theta-\beta)}{\int_{\mathbb{R}} f(\alpha\theta-\beta)G(d\theta)}.$$

Thus,

$$\frac{\partial}{\partial \alpha}F\big[\alpha\theta - \overline{p}(\alpha,\gamma)\big] = \left\{\theta - \frac{\partial}{\partial\alpha}\overline{p}(\alpha,\gamma)\right\} f\big[\alpha\theta - \overline{p}(\alpha,\gamma)\big], \tag{B.2}$$

where $\frac{\partial}{\partial\alpha}\overline{p}[\alpha,\gamma] = E_{\alpha,\overline{p}[\alpha,\gamma]}[X]$.

Let

$$\begin{aligned}\gamma_{jk} &= P[Y_{ij}=1, Y_{ik}=1 \mid \boldsymbol{\alpha}_{1:J}, \boldsymbol{\beta}_{1:J}]\\ &= \int_{\mathbb{R}} F[\alpha_j\theta-\beta_j]F[\alpha_k\theta-\beta_k]G(d\theta) \doteq g(\alpha_j,\beta_j,\alpha_k,\beta_k)\\ &= \int_{\mathbb{R}} F\big[\alpha_j\theta-\overline{p}(\alpha_j,\gamma_j)\big]F\big[\alpha_k\theta-\overline{p}(\alpha_k,\gamma_k)\big]G(d\theta) \doteq h(\alpha_j,\gamma_j,\alpha_k,\gamma_k).\end{aligned} \tag{B.3}$$

Using (B.2), it follows that

$$\begin{aligned}&\frac{\partial}{\partial\alpha_j}h(\alpha_j,\gamma_j,\alpha_k,\gamma_k)\\ &\quad= \int_{\mathbb{R}} f\big[\alpha_j\theta-\overline{p}(\alpha_j,\gamma_j)\big]G(d\theta) \times C_{\alpha_j,\overline{p}(\alpha_j,\gamma_j)}\big\{X, F\big[\alpha_k X - \overline{p}(\alpha_k,\gamma_k)\big]\big\} > 0\end{aligned} \tag{B.4}$$

provided $\alpha_k > 0$ since in this case $F[\alpha_k X - \overline{p}(\alpha_k,\gamma_k)]$ is a strictly increasing function of $X$ and, consequently, the covariance between $X$ and $F[\alpha_k X - \overline{p}(\alpha_k,\gamma_k)]$ is positive (if $X$ is not degenerate). Thus, $h(\alpha_j,\gamma_j,\alpha_k,\gamma_k)$ is a strictly increasing function in $\alpha_j$. Similarly, it is concluded that $h(\alpha_j,\gamma_j,\alpha_k,\gamma_k)$ is also a strictly increasing function in $\alpha_k$ provided $\alpha_j > 0$. The inverse function of $h$ can, therefore, be defined as

$$\overline{h}(\alpha_j,\gamma_j,\zeta,\gamma_k) = \inf\big\{\alpha'_k : h\big(\alpha_j,\gamma_j,\alpha'_k,\gamma_k\big) > \zeta\big\}$$

and consequently,

$$\begin{aligned}&\text{(i)}\quad \overline{h}\big[\alpha_j,\gamma_j,h(\alpha_j,\gamma_j,\alpha_k,\gamma_k),\gamma_k\big] = \alpha_k;\\ &\text{(ii)}\quad h\big[\alpha_j,\gamma_j,\overline{h}(\alpha_j,\gamma_j,\zeta_k,\gamma_k),\gamma_k\big] = \zeta.\end{aligned} \tag{B.5}$$

### B.3. Identification of the Discrimination Parameters

Let $J \geq 3$. Using (B.3), we have that

$$\gamma_{12} = h(\alpha_1, \gamma_1, \alpha_2, \gamma_2), \qquad \gamma_{13} = h(\alpha_1, \gamma_1, \alpha_3, \gamma_3), \qquad \gamma_{23} = h(\alpha_2, \gamma_2, \alpha_3, \gamma_3).$$

Therefore, by (B.5.i) it follows that $\alpha_2 = \overline{h}(\alpha_1, \gamma_1, \gamma_{12}, \gamma_2)$ and $\alpha_3 = \overline{h}(\alpha_1, \gamma_1, \gamma_{13}, \gamma_3)$. Thus,

$$\begin{aligned} \gamma_{23} &= h\big[\overline{h}(\alpha_1, \gamma_1, \gamma_{12}, \gamma_2), \gamma_2, \overline{h}(\alpha_1, \gamma_1, \gamma_{13}, \gamma_3), \gamma_3\big] \\ &\doteq k(\alpha_1, \gamma_1, \gamma_2, \gamma_3, \gamma_{12}, \gamma_{13}). \end{aligned}$$

The identification of $\alpha_1$ follows because the function $k$ is invertible. As a matter of fact,

$$\begin{aligned} &\frac{\partial}{\partial \alpha_1} k(\alpha_1, \gamma_1, \gamma_2, \gamma_3, \gamma_{12}, \gamma_{13}) \\ &\quad = \frac{\partial h}{\partial \alpha_1}\big[\overline{h}(\alpha_1, \gamma_1, \gamma_{12}, \gamma_2), \gamma_2, \overline{h}(\alpha_1, \gamma_1, \gamma_{13}, \gamma_3), \gamma_3\big] \times \frac{\partial \overline{h}}{\partial \alpha_1}[\alpha_1, \gamma_1, \gamma_{12}, \gamma_2] \\ &\qquad + \frac{\partial h}{\partial \alpha_2}\big[\overline{h}(\alpha_1, \gamma_1, \gamma_{12}, \gamma_2), \gamma_2, \overline{h}(\alpha_1, \gamma_1, \gamma_{13}, \gamma_3), \gamma_3\big] \times \frac{\partial \overline{h}}{\partial \alpha_1}[\alpha_1, \gamma_1, \gamma_{12}, \gamma_2]. \end{aligned}$$

But

$$\frac{\partial \overline{h}}{\partial \alpha_1}[\alpha_1, \gamma_1, \gamma_{12}, \gamma_2] = -\frac{\frac{\partial h}{\partial \alpha_i}[\alpha_1, \gamma_1, \overline{h}(\alpha_1, \gamma_1, \gamma_{12}, \gamma_2), \gamma_2]}{\frac{\partial h}{\partial \alpha_2}[\alpha_1, \gamma_1, \overline{h}(\alpha_1, \gamma_1, \gamma_{12}, \gamma_2), \gamma_2]}.$$

Using (B.4), we conclude that $\frac{\partial}{\partial \alpha_1} k(\alpha_1, \gamma_1, \gamma_2, \gamma_3, \gamma_{12}, \gamma_{13}) < 0$ and, therefore, $k$ is invertible. Finally, by (B.5.i), the identification of the remaining discrimination parameters then follows.

The previous arguments have been established assuming that the distribution generating the person specific abilities is known. If such a distribution is known up to the scale parameter $\sigma$, the previous arguments apply for $\widetilde{\alpha}_j = \alpha_j \sigma$. Thus, we obtain the following theorem.

**Theorem B.1.** *Consider the statistical model induced by both the* 2*PL model* (B.1) *and the person specific abilities distributed according to a distribution* $G$ *known up to the scale parameter* $\sigma$, *where the* $F$ *is a strictly continuous increasing distribution function with a density function* $f$ *strictly positive on* $\mathbb{R}$. *The parameters of interest* $(\boldsymbol{\alpha}_{1:J}, \boldsymbol{\beta}_{1:J}, \sigma)$ *are identified by one observation provided that*

1. *At least three items are available.*
2. *The discrimination parameter* $\alpha_1$ *is fixed at* 1.

*If the distribution* $G$ *is fully known, then parameters of interest* $(\boldsymbol{\alpha}_{1:J}, \boldsymbol{\beta}_{1:J})$ *are identified provided that at least three items are available.*

It is relevant to remark that the positivity of the discrimination parameters is established by the identification analysis.

## Appendix C. Proof of Theorem 5

The identification analysis of the parameters of interest $(\boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G)$ should be done in the asymptotic Bayesian model defined on $(\boldsymbol{Y}_i, \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G)$, where $\boldsymbol{Y}_i \in \{0, 1\}^{\mathbb{N}}$ corresponds to the response pattern of person $i$. According to Definition 2, the corresponding minimal sufficient parameter is given by the following $\sigma$-field:

$$\mathcal{A}^* \doteq \sigma\big\{E(f \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G) : f \in \big[\sigma(\boldsymbol{Y}_1)\big]^+\big\},$$

where $[\sigma(\boldsymbol{Y}_1)]^+$ denotes the set of positive functions $f$ such that $f = g(\boldsymbol{Y}_1)$, with $g$ a measurable function. The identification of the semi-parametric 1PL-G model leads to proving, under identification restrictions if necessary, that $(\boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G)$ is a measurable function of the parameter $\mathcal{A}^*$. By the Doob–Dynkin lemma, this is equivalent to proving that, under identification restrictions if necessary,

$$\sigma(\boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G) = \mathcal{A}^*.$$

This equality relies on the following steps:

STEP 1: By the same arguments used to establish identity (4.4), it follows that

$$\sigma(\boldsymbol{\beta}_{2:\infty}, \boldsymbol{\delta}_{2:\infty}) \doteq \sigma(\beta_j : j \geq 2) \vee \sigma(\delta_j : j \geq 2) \subset \mathcal{A}^*,$$

where $\delta_j \doteq 1 - c_j$.

STEP 2: Hypotheses **H1**, **H2**, **H3**, and **H4** jointly imply that $\{(u_j, v_j) : 2 \leq j < \infty\}$ are *iid* conditionally on $(\beta_1, \delta_1, K, H)$. By the Strong Law of Large Numbers, it follows that

$$W^{\beta_1,\delta_1}(B) \doteq P\big[(u_2, v_2) \in B \mid \beta_1, \delta_1, K, H\big] \overset{\text{a.s.}}{=} \limsup_m \frac{1}{m} \sum_{1 \leq j \leq m} \mathbb{1}_{\{(u_j,v_j) \in B\}}$$

for $B \in \mathcal{B}^+ \times \mathcal{B}$. But Propositions 2 and 3 ensure that $\{u_j : 2 \leq j < \infty\}$ and $\{v_j : 2 \leq j < \infty\}$ are identified parameters. It follows that $\{(u_j, v_j) : 2 \leq j < \infty\}$ is measurable w.r.t. $\mathcal{A}^*$ and, consequently, $W^{\beta_1,\delta_1}(B)$ is measurable w.r.t. $\overline{\mathcal{A}^*}$ for all $B \in \mathcal{B}^+ \times \mathcal{B}$. The upper-bar denotes a $\sigma$-field completed with measurable sets; for a definition, see Chapter 2 in Florens et al. (1990).

STEP 3: Using (4.4), it follows that

$$\begin{aligned} E(Y_{ij} \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, \boldsymbol{\theta}_{1:\infty}) &= 1 - \frac{\delta_j e^{\beta_j}}{e^{\beta_j} + e^{\theta_i}} \\ &= 1 - \frac{u_j}{v_j + \frac{1}{\delta_1} + \frac{1}{\delta_1 \exp(\beta_1)} \exp(\theta_i)}, \end{aligned}$$

and

$$\begin{aligned} \overline{p}_{iJ} &\doteq E[\overline{Y}_{iJ} \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, \boldsymbol{\theta}_{1:\infty}] \\ &= 1 - \frac{1}{J} \sum_{1 \leq j \leq J} \frac{u_j}{v_j + \frac{1}{\delta_1} + \frac{1}{\delta_1 \exp(\beta_1)} \exp(\theta_i)}. \end{aligned}$$

STEP 3.A: By the law of large deviations (see Shiryaev, 1995, Chapter IV, Section 5), it follows that

$$\overline{Y}_{iJ} - \overline{p}_{iJ} \longrightarrow 0 \quad \text{a.s. conditionally on } (\beta_1, \delta_1, \theta_i) \text{ as } J \to \infty.$$

But as $J \to \infty$

$$\begin{aligned}
\overline{p}_{iJ} &\longrightarrow 1 - E\left[\frac{u_j}{v_j + \frac{1}{\delta_1} + \frac{1}{\delta_1 \exp(\beta_1)} \exp(\theta_i)} \,\Big|\, \beta_1, \delta_1, \theta_i\right] \quad \text{a.s. conditionally on}(\beta_1, \delta_1, \theta_i) \\
&= \int_{\mathbb{R}_+ \times \mathbb{R}} \left\{1 - \frac{u}{v + \frac{1}{\delta_1} + \frac{1}{\delta_1 \exp(\beta_1)} \exp(\theta_i)}\right\} W^{\beta_1, \delta_1}(du, dv) \\
&\doteq p(\beta_1, \delta_1, \theta_i).
\end{aligned}$$

Therefore,

$$\overline{Y}_{iJ} \longrightarrow p(\beta_1, \delta_1, \theta_i) \quad \text{a.s. conditionally on } (\beta_1, \delta_1, \theta_i) \text{ as } J \to \infty.$$

STEP 3.B: It follows that, for all $g \in C_b([0, 1])$,

$$g(\overline{Y}_{iJ}) \longrightarrow g\big(p(\beta_1, \delta_1, \theta_i)\big) \quad \text{a.s. and in } L^1 \text{ conditionally on } (\beta_1, \delta_1, \theta_i) \text{ as } J \to \infty.$$

Then for all $g \in C_b([0, 1])$,

$$\begin{aligned}
E\big[g(\overline{Y}_{iJ}) \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G\big] &\longrightarrow E\big[g\big(p(\beta_1, \delta_1, \theta_i)\big) \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G\big] \quad \text{a.s.} \\
&= \int g\left\{1 - E\left[\frac{u_j}{v_j + \frac{1}{\delta_1} + \frac{1}{\delta_1 \exp(\beta_1)} \exp(\theta_i)} \mid \beta_1, \delta_1, \theta_i\right]\right\} G(d\theta).
\end{aligned}$$

By definition of conditional expectation, $E[g(\overline{Y}_{iJ}) \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G]$ is measurable w.r.t. $\mathcal{A}^*$. Thus,

$$\int_{\mathbb{R}_+} g\left\{1 - E\left[\frac{u_j}{v_j + \frac{1}{\delta_1} + \frac{1}{\delta_1 \exp(\beta_1)} \exp(\theta_i)} \,\Big|\, \beta_1, \delta_1, \theta_i\right]\right\} G(d\theta)$$

is measurable w.r.t. $\overline{\mathcal{A}^*}$; the bar is added because such an integral is the a.s. limit of the sequence $\{E[g(\overline{Y}_{iJ}) \mid \boldsymbol{\beta}_{1:\infty}, \boldsymbol{c}_{1:\infty}, G] : J \in \mathbb{N}\}$.

STEP 3.C: Using the transformation (4.10), it is follows that

$$\int_{\mathbb{R}_+} g\big[L(x)\big] G_{\beta_1, \delta_1}(dx) \quad \text{is } \overline{\mathcal{A}^*}\text{-measurable},$$

where

$$L(x) = \int_{\mathbb{R}_+ \times \mathbb{R}} \left(1 - \frac{u}{v + x}\right) W^{\beta_1, \delta_1}(du, dv).$$

The function $L(\cdot)$ is a strictly continuous function from $(\delta^{-1}, \infty)$ to $(0, 1)$ that is known because it is measurable w.r.t. $\sigma(W^{\beta_1, \delta_1})$. By STEP 2, $\sigma(W^{\beta_1, \delta_1}) \subset \overline{\mathcal{A}^*}$. In particular, for every function $f \in C_b(\mathbb{R}^+)$, take $g(y) = f[\overline{L}(y)]$, where $\overline{L}(\alpha) = \inf\{x : L(x) \geq \alpha\}$. It follows that

$$\int_{\mathbb{R}_+} f(x) G_{\beta_1, \delta_1}(dx)$$

is measurable w.r.t. $\overline{\mathcal{A}^*}$. Considering

$$f_n(y) = \mathbb{1}_{(0,x]}(y) + \big[1 - n(y - x)\big] \mathbb{1}_{(x, x + \frac{1}{n})}(y) \downarrow \mathbb{1}_{(0,x]}(y) \quad \forall x \in \mathbb{R}^+,$$

as $n \to \infty$, the monotone convergence theorem implies that, for every $x \in \mathbb{R}^+$, $G_{\beta_1, \delta_1}((0, x])$, and so $G_{\beta_1, \delta_1}$, is measurable w.r.t. $\overline{\mathcal{A}^*}$.

STEP 4: From STEPS 1 and 3C, it follows that $(\boldsymbol{\beta}_{2:\infty}, \boldsymbol{\delta}_{2:\infty}, G_{\beta_1,\delta_1})$ is measurable w.r.t. $\overline{\mathcal{A}^*}$. By the Doob–Dynkin lemma, this is equivalent to

$$\sigma(\boldsymbol{\beta}_{2:\infty}, \boldsymbol{\delta}_{2:\infty}) \vee \sigma(G_{\beta_1,\delta_1}) \subset \overline{\mathcal{A}^*}.$$

However, $\sigma(G_{\beta_1,\delta_1}) \subset \sigma(G) \vee \sigma(\beta_1, \delta_1)$. Therefore, two restrictions should be introduced in order to obtain the equality $\sigma(G_{\beta_1,\delta_1}) = \sigma(G) \vee \sigma(\beta_1, \delta_1)$. Two possibilities can be considered:

1. The first possibility consists in fixing two $q$-quantiles of $G$. In fact, let

$$x_1 = \inf\{x : G_{\beta_1,\delta_1}(x) > q_1\}, \qquad x_2 = \inf\{x : G_{\beta_1,\delta_1}(x) > q_2\}.$$

Using (4.10), this is equivalent to

$$\beta_1 + \ln(\delta_1 x_1 - 1) = y_1 = \inf\{y : G(y) > q_1\},$$
$$\beta_1 + \ln(\delta_1 x_2 - 1) = y_2 = \inf\{y : G(y) > q_2\}.$$

It follows that

$$\beta_1 = \ln\left[\frac{x_1 e^{y_2} - x_2 e^{y_1}}{x_2 - x_1}\right], \qquad \delta_1 = \frac{e^{y_2} - e^{y_1}}{x_1 e^{y_2} - x_2 e^{y_1}};$$

that is, $\beta_1$ and $\delta_1$ are identified since $x_1, x_2, y_1, y_2$ depends on $G$ that it is identified.

2. The second possibility consists in fixing the mean and the variance of the distribution of $\exp(\theta)$, namely

$$E_G(e^\theta) = \mu, \qquad V_G(e^\theta) = \sigma^2.$$

Using (4.10), this is equivalent to

$$m = E_{G_{\beta_1,\delta_1}}(X) = \frac{1}{\delta_1} + \frac{\mu}{\delta_1 e^{\beta_1}}, \qquad v^2 = V_{G_{\beta_1,\delta_1}}(X) = \frac{\sigma^2}{\delta_1^2 e^{2\beta_1}}.$$

It follows that

$$\beta_1 = \ln\left(\frac{m\sigma}{v} - \mu\right), \qquad \delta_1^{-1} = m - \frac{\mu v}{\sigma}.$$

For instance, if $\mu = 0$ and $\sigma = 1$, then

$$\beta_1 = \ln\left(\frac{m}{v}\right), \qquad \delta_1 = \frac{1}{m};$$

that is, $\beta_1$ and $\delta_1$ are identified since $m$ and $v$ depend on $G$ which is identified.

## References


Adams, R.J., Wilson, M.R., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*, 1–23.

Adams, R.J., & Wu, M.L. (2007). The mixed-coefficients multinomial logit model: a generalization form of the Rasch model. In M. von Davier & C.H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (pp. 57–75). Berlin: Springer.

Andersen, E.B. (1980). *Discrete statistical models with social sciences applications*. Amsterdam: North-Holland.

Béguin, A.A., & Glas, C.A.W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*, 541–562.

Berti, P., Pratelli, L., & Riggo, P. (2008). Trivial intersection of $\sigma$-fields and Gibbs sampling. *Annals of Probability*, *36*, 2215–2234.

Berti, P., Pratelli, L., & Riggo, P. (2010). Atomic intersection of $\sigma$-fields and some of its consequences. *Probability Theory and Related Fields*, *148*, 269–283.
Birnbaum, A. (1968). Some latent trait models and their use in inferring any examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Reading: Addison-Wesley.
Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, *46*, 443–459.
Bock, R.D., & Zimowski, M.F. (1997). Multiple group IRT. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 433–448). Berlin: Springer.
Carlin, B.P., & Louis, T.A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). London: Chapman & Hall/CRC.
De Boeck, P., & Wilson, M. (2004). *Explanatory item response models. A generalized linear and nonlinear approach*. Berlin: Springer.
Del Pino, G., San Martín, E., González, J., & De Boeck, P. (2008). On the relationships between sum score based estimation and joint maximum likelihood estimation. *Psychometrika*, *73*, 145–151.
Eberly, E.E., & Carlin, B.P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial data. *Statistics in Medicine*, *19*, 2279–2294.
Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum Associates.
Florens, J.-P., & Mouchart, M. (1982). A note on noncausality. *Econometrica*, *50*, 583–591.
Florens, J.-P., Mouchart, M., & Rolin, J.-M. (1990). *Elements of Bayesian statistics*. New York: Dekker.
Florens, J.-P., & Rolin, J.-M. (1984). Asymptotic sufficiency and exact estimability. In J.-P. Florens, M. Mouchart, J.-P. Raoult, & L. Simar (Eds.), *Alternative approaches to time series analysis* (pp. 121–142). Bruxelles: Publications des Facultés Universitaires Saint-Louis.
Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, *8*, 261–263.
Gelfand, A.E., & Sahu, S.K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, *94*, 247–253.
Gelman, A., & Rubin, R. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.
Geweke, J. (1992). *Bayesian statistics: Vol. 4. Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. Oxford: Clarendon Press.
Ghosh, M., Ghosh, A., Chen, M.-H., & Agresti, A. (2000). Noninformative 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 Sciences*, *20*, 111–140.
Halmos, P. (1951). *Introduction to Hilbert space, and the theory of spectral multiplicity*. New York: Chelsea.
Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Thousand Oaks: Sage.
Heidelberg, P., & Welch, P. (1992). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*, 1109–1144.
Hutschinson, T.P. (1991). *Ability, parameter information, guessing: statistical modelling applied to multiple-choice tests*. Rundle Mall: Rumsby Scientific Publishing.
Karabatsos, G., & Walker, S. (2009). Coherent psychometric modelling with Bayesian nonparametrics. *British Journal of Mathematical & Statistical Psychology*, *62*, 1–20.
Kass, R., Carlin, B., Gelman, A., & Neal, R. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *American Statistician*, *52*, 93–100.
Koopmans, T.C., & 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.
van der Linden, W., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. Berlin: Springer.
Lindley, D.V. (1971). *Bayesin statistics: a review*. Philadelphia: Society for Industrial and Applied Mathematics.
Maris, G., & Bechger, T. (2009). On interpreting the model parameters for the three parameter logistic model. *Measurement Interdisciplinary Research & Perspective*, *7*, 75–86.
McDonald, R.P. (1999). *Test theory: a unified treatment*. Hillsdale: Erlbaum.
Millsap, R., & Maydeu-Olivares, A. (2009). *Quantitative methods in psychology*. Thousand Oaks: Sage.
Miyazaki, K., & Hoshino, T. (2009). A Bayesian semiparametric item response model with Dirichlet process priors. *Psychometrika*, *74*, 375–393.
Molenaar, I.W. (1995). Estimation of item parameters. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models. Foundations, recent developments and applications*. New York: Springer (Chapter 3).
Mouchart, M. (1976). A note on Bayes theorem. *Statistica*, *36*, 349–357.
Poirier, D.J. (1998). Revising beliefs in nonidentified models. *Econometric Theory*, *14*, 483–509.
R Development Core Team (2006). *R: a language and environment for statistical computing* [Computer software manual]. Vienna, Austria. http://www.R-project.org (ISBN 3-900051-07-0).
Rao, M.M. (2005). *Conditional measures and applications* (2nd ed.). London: Chapman & Hall/CRC.
Rizopoulos, D. (2006). ltm: an R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1–25. http://www.jstatsoft.org/v17/i05/.
Roberts, G.O., & Rosenthal, J. (1998). Markov-chain Monte Carlo: some practical implications of theoretical results. *Canadian Journal of Statistics*, *26*(1), 5–20.

San Martín, E., Del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, *30*, 183–203.
San Martín, E., & González, J. (2010). Bayesian identifiability: contributions to an inconclusive debate. *Chilean Journal of Statistics*, *1*, 69–91.
San Martín, E., González, J., & Tuerlinckx, F. (2009). Identified parameters, parameters of interest and their relationships. *Measurement Interdisciplinary Research & Perspective*, *7*, 95–103.
San Martín, E., Jara, A., Rolin, J.-M., & Mouchart, M. (2011). On the Bayesian nonparametric generalization of IRT-type models. *Psychometrika*, *76*, 385–409.
San Martín, E., Mouchart, M., & Rolin, J.M. (2005). Ignorable common information, null sets and Basu's first theorem. *Sankhyā*, *67*, 674–698.
San Martín, E., & Quintana, F. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, *16*, 99–106.
Shiryaev, A.N. (1995). *Probability* (2nd ed.). Berlin: Springer.
Spivak, M. (1965). *Calculus on manifolds: a modern approach to classical theorems of advanced calculus*. Cambridge: Perseus Book Publishing.
Swaminathan, H., & Gifford, J.A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika*, *51*, 589–601.
Thissen, D. (2009). On interpreting the parameters for any item response model. *Measurement Interdisciplinary Research & Perspective*, *7*, 104–108.
Thissen, D., & Wainer, H. (2001). *Item response models for items scored in two categories*. Berlin: Springer.
Woods, C.M. (2006). Ramsay-curve item response theory (RC-IRT) to detect and correct for nonnormal latent variables. *Psychological Methods*, *11*, 253–270.
Woods, C.M. (2008). Ramsay-curve item response theory for the three-parameter logistic item response model. *Applied Psychological Measurement*, *32*, 447–465.
Woods, C.M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, *71*, 281–301.
Xie, Y., & Carlin, B.P. (2006). Measure of Bayesian learning and identifiability in hierarchical models. *Journal of Statistcal Planning and Inference*, *136*, 3458–3477.

*Manuscript Received: 26 SEP 2011*
*Final Version Received: 17 APR 2012*
*Published Online Date: 1 FEB 2013*