PSYCHOMETRIKA
2012
DOI: 10.1007/S11336-012-9253-9

# IDENTIFICATION OF A SEMIPARAMETRIC ITEM RESPONSE MODEL

MICHAEL PERESS

UNIVERSITY OF ROCHESTER

We consider the identification of a semiparametric multidimensional fixed effects item response model. Item response models are typically estimated under parametric assumptions about the shape of the item characteristic curves (ICCs), and existing results suggest difficulties in recovering the distribution of individual characteristics under nonparametric assumptions. We show that if the shape of the ICCs are unrestricted, but the shape is common across individuals and items, the individual characteristics are identified. If the shape of the ICCs are allowed to differ over items, the individual characteristics are identified in the multidimensional linear compensatory case but only identified up to a monotonic transformation in the unidimensional case. Our results suggest the development of two new semiparametric estimators for the item response model.

Key words: item response theory, nonparametric identification.

## 1. Introduction

The item response framework has applications in psychometrics, political science, and marketing. In all cases, the goal is to recover latent individual characteristics in a multidimensional space from binary response data. These characteristics include ability (in the psychometrics literature), ideal points (in the political science literature), and product characteristics (in the marketing literature). These responses include the correct response to a test item (in the psychometrics literature), the choice of which of two options to vote for (in the political science literature), and the choice of which product to purchase (in the marketing literature). Item response estimators are useful in recovering a multidimensional latent space from a sequence of individual responses.

Item response models are typically estimated under parametric assumptions about the shape of the item characteristic curves (ICCs), using random effects estimators (Bock & Lieberman, 1970; Bock & Aitken, 1981) or fixed effects estimators (Lord, 1980; Poole & Rosenthal, 1991, 1997). Using parametric estimators, we can recover the individual traits on an interval scale. Nonparametric random effects estimators based on recovering ICCs have been applied to psychometric data (Ramsay & Abrahamowicz, 1989; Ramsay, 1991; Douglas, 1997; Sijtsma, 1998), but the shape of the distribution of individual characteristics is not identified (Douglas, 2001). Similarly, Poole's (2000) nonparametric fixed effects estimator can only recover the ordering of the individual characteristics in one dimension. This suggests that while percentile ranks of individual traits can be recovered under a weak set of assumptions, the shape of the distribution of individual characteristics recovered by parametric estimators such as the normal ogive model is potentially an artifact of arbitrary and untestable distributional and functional form assumptions.

Interval scale estimates of individual characteristics are useful for a number of reasons. Estimates of individual traits from test scores may enter into the decision process of a college admissions committee. The committee may desire to enter these scores into a linear additive model. In

**Electronic Supplementary Material** The online version of this article (doi:10.1007/s11336-012-9253-9) contains supplementary material, which is available to authorized users.

Requests for reprints should be sent to Michael Peress, Department of Political Science, University of Rochester, Rochester, NY, USA. E-mail: mperess@mail.rochester.edu



Published online: 08 February 2012

the absence of interval scale estimates of these traits, however, the committee may need to consider non-linear transformations to map the percentile ranks onto an appropriate scale. In many applications, we would like to link groups of individuals who have responded to separate sets of items. For example, when administering a standardized test, it is desirable to change the items over time to prevent test-takers from gaining prior knowledge of the items while maintaining comparability across different tests. If interval scale estimates are identified, a small number of bridge items that are common across successive tests will allow for comparability. Achieving this comparability using ordinal (or percentile) ratings of individuals is far more demanding.

We study the identification of a semiparametric fixed effects item response model. We show that this model is (asymptotically) identified under reasonable conditions. Our results suggest that interval scale estimates of individual characteristics are not artifacts of distributional assumptions (at least in the fixed effects case). Instead, if the shapes of the ICCs are unrestricted, but common across individuals and items, the individual characteristics are identified. If the shapes of the ICCs are allowed to differ over items, the individual characteristics are identified in the multidimensional linear compensatory case. It is only in the unidimensional case where the shapes of the ICCs differ across items where we encounter identification problems. In this case, we can only identify the individual characteristics up to a monotonic transformation (or equivalently, on an ordinal scale). Our results suggest that the development of semiparametric item response estimators that produce interval level estimates of individual characteristics is possible and identification of individual characteristics fails only under unfavorable assumptions.

## 2. Literature

### 2.1. Random Effects Estimators

Parametric random effects item response models often consider the following linear compensatory model for the data:

$$\begin{aligned}&\Pr(y_{n1}, y_{n2}, \ldots, y_{nT}; a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\xi})\\&\quad= \int_{\boldsymbol{\theta}} \left\{ \prod_{t=1}^{T} F\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}\big)^{y_{nt}} \big[1 - F\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}\big)\big]^{1-y_{nt}} \right\} h(\boldsymbol{\theta}; \boldsymbol{\xi})\, d\boldsymbol{\theta}. \qquad (1)\end{aligned}$$

Here, $y_{nt} = 1$ denotes a correct response by individual $n$ to item $t$ and $y_{nt} = 0$ denotes an incorrect response. We assume that the individual characteristics, $\boldsymbol{\theta}_n$, satisfy $\boldsymbol{\theta}_n \in \mathbb{R}^D$, and the item characteristics, $(a_t, \boldsymbol{b}_t)$, satisfy $a_t \in \mathbb{R}$ and $\boldsymbol{b}_t \in \mathbb{R}^D$. $D \in \mathbb{N}$ denotes the dimensionality of individual characteristics and $F$ is assumed to be known. $F$ is often taken to be standard normal (as in the normal ogive model) or logistic (as in Birnbaum's two-parameter logistic model) and $h(\boldsymbol{\theta}; \boldsymbol{\xi})$ is known up to the parameter vector $\boldsymbol{\xi}$ and is often taken to be multivariate normal with $h(\boldsymbol{\theta}; \boldsymbol{\xi}) = \phi(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Omega})$. Parametric random effects models impose parametric assumptions on $F$ (which governs the shapes of the ICCs) and $h$ (the distribution of individual characteristics).

Item response models can be thought of as latent variable models with

$$y_{nt}^* = a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n - \varepsilon_{nt}. \qquad (2)$$

Here, $y_{nt} = 1\{y_{nt}^* \geq 0\}$, and $\varepsilon_{nt}$ are independent errors with $\varepsilon_{nt} \sim F$. Hence, we can refer to $F$ as the distribution of the error term. We normalize $F$ to satisfy $F(0) = \frac{1}{2}$ and $F(1) = p > \frac{1}{2}$. Treating $\boldsymbol{\theta}_n$ as a random effect with distribution $h(\boldsymbol{\theta}; \boldsymbol{\xi})$ leads to the same formulation as (1).

Nonparametric random effects models do not impose parametric assumptions on $F$ and $h$. Instead, we have

$$\begin{aligned}&\Pr(y_{n1}, y_{n2}, \ldots, y_{nT}; a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, F, h)\\&\quad = \int_{\boldsymbol{\theta}} \left\{ \prod_{t=1}^{T} F\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}\big)^{y_{nt}} \big[1 - F\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}\big)\big]^{1-y_{nt}} \right\} h(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \qquad (3)\end{aligned}$$

A less restrictive model allows $F$ to differ across items,

$$\begin{aligned}&\Pr(y_{n1}, y_{n2}, \ldots, y_{nT}; a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, F_1, \ldots, F_T, h)\\&\quad = \int_{\boldsymbol{\theta}} \left\{ \prod_{t=1}^{T} F_t\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}\big)^{y_{nt}} \big[1 - F_t\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}\big)\big]^{1-y_{nt}} \right\} h(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \qquad (4)\end{aligned}$$

An even more general model drops the linear compensatory assumption,

$$\begin{aligned}&\Pr(y_{n1}, y_{n2}, \ldots, y_{nT}; G_1, \ldots, G_T, h)\\&\quad = \int_{\boldsymbol{\theta}} \left\{ \prod_{t=1}^{T} G_t(\boldsymbol{\theta})^{y_{nt}} \big[1 - G_t(\boldsymbol{\theta})\big]^{1-y_{nt}} \right\} h(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \qquad (5)\end{aligned}$$

Here, the functions $G_t$ are referred to as the ICCs. The distinction between (4) and (5) is only relevant in the multidimensional case. In the unidimensional case, model (5) is a reparametrization of (4) where

$$G_t(\theta) = F_t(a_t + b_t\theta). \qquad (6)$$

The model specified in (5) when $D = 1$ is considered in Douglas (2001). Douglas shows that this model is not identified. We say that model (5) is identified if there does not exist a $(G_1, \ldots, G_T, h) \neq (G_{10}, \ldots, G_{T0}, h_0)$ such that

$$\begin{aligned}&\int_{\theta} \left\{ \prod_{t=1}^{T} G_t(\theta)^{y_{nt}} \big[1 - G_t(\theta)\big]^{1-y_{nt}} \right\} h(\theta)\, d\theta\\&\quad = \int_{\theta} \left\{ \prod_{t=1}^{T} G_{t0}(\theta)^{y_{nt}} \big[1 - G_{t0}(\theta)\big]^{1-y_{nt}} \right\} h_0(\theta)\, d\theta \quad \text{for all } n. \qquad (7)\end{aligned}$$

Let $\circ$ indicate the composition of two functions, i.e., $[f_1 \circ f_2](x) = f_1(f_2(x))$. Consider any $(G_1, \ldots, G_T, h)$ satisfying $G_t = G_{t0} \circ H$ where $H(\theta) = \int_{u=-\infty}^{\theta} h(u)\, du$. We have

$$\begin{aligned}&\int_{\theta} \left\{ \prod_{t=1}^{T} G_t(\theta)^{y_{nt}} \big[1 - G_t(\theta)\big]^{1-y_{nt}} \right\} h(\theta)\, d\theta\\&\quad = \int_{\theta} \left\{ \prod_{t=1}^{T} G_{t0}\big(H(\theta)\big)^{y_{nt}} \big[1 - G_{t0}\big(H(\theta)\big)\big]^{1-y_{nt}} \right\} h(\theta)\, d\theta\\&\quad = \int_{\theta} \left\{ \prod_{t=1}^{T} G_{t0}(\theta)^{y_{nt}} \big[1 - G_{t0}(\theta)\big]^{1-y_{nt}} \right\} d\theta. \qquad (8)\end{aligned}$$

The last equality follows from the application of the transformation, $\delta = H(\theta)$. The result implies that any density of individual characteristics with full support is consistent with the observable data. In light of this result, nonparametric estimators based on the ICCs assume a Uniform$(0, 1)$ distribution for $\theta$. Douglas (2001) shows that once this restriction is imposed, the item response model is asymptotically nonparametrically identified. Hence, these estimators can recover the ordering of the individual characteristics, but the shape of the distribution of individual characteristics produced in the second stage of the estimation is an artifact of this assumption.

The model described by (4) and (5) may be more general than necessary. The model treats items and individuals asymmetrically—the model implicitly assumes that the shape of the ICCs (or equivalently, the distribution of the error term) is common across individuals, but differs across items. In particular, it may be reasonable to assume that the form of $F_t$ is common across items, as it is already being assumed to be common across individuals. Second, the linear compensatory framework may be sufficiently general for some applications. We provide identification results which address both of these possibilities.

### 2.2. *Fixed Effects Estimators*

Under an alternative framework, which we refer to as the "fixed effects model", we have

$$\Pr(y_{nt}; \boldsymbol{\theta}_n, a_t, \boldsymbol{b}_t, F) = F\left(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n\right)^{y_{nt}}\left[1 - F\left(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n\right)\right]^{1-y_{nt}}. \tag{9}$$

The term fixed effects model has no widely accepted definition, but we use it here to refer to the types of item response model that would typically be estimated using joint maximum likelihood. We reserve the term "random effects model" to refer to those item response models which would typically be estimated using marginal maximum likelihood. If $F$ is the logistic cdf, then we have Birnbaum's (1968) two-parameter logistic model. If $F$ has an unknown functional form, then we have a semiparametric linear compensatory fixed effects model. More generally, we may allow $F$ to differ across items,

$$\Pr(y_{nt}; \boldsymbol{\theta}_n, a_t, \boldsymbol{b}_t, F_t) = F_t\left(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n\right)^{y_{nt}}\left[1 - F_t\left(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n\right)\right]^{1-y_{nt}}. \tag{10}$$

Poole (2000) develops a nonparametric estimator for the fixed effects item response model based on optimal classification. The estimator is only able to recover the ordering of individual characteristics in the unidimensional case, so once again, interval scale information is not forthcoming from the procedure. This has led to speculation that the non-identification is not simply a property of Poole's estimator, but is a property of the nonparametric item response problem. This result, if true, would be of consequence to three sets of literature—psychometric, political science, and marketing. It is this fact that we examine in detail in this paper.

## 3. Identification

### 3.1. *Assumptions*

We work within the fixed effects framework. We distinguish between the homogeneous (9) and heterogeneous (10) versions of the model. We assume that (9) and (10) hold for $n \in \mathcal{N}$ and $t \in \mathcal{T}$. If $\mathcal{N}$ and $\mathcal{T}$ are finite sets with $N = |\mathcal{N}|$ and $T = |\mathcal{T}|$, it is clear that the model will not be identified since $F$ is an infinite dimensional quantity, but we only observe $NT$ quantities in the data. Our identification results therefore rely on $\mathcal{N} = \mathbb{N}$ and $\mathcal{T} = \mathbb{N}$. This approach is reasonable since only large sample identification is typically needed in a proof of consistency. A similar focus on asymptotic identification is used by Douglas (2001).

Throughout, we use zero subscripts to denote parameters of the data generating process and we let $\|\boldsymbol{x}\|$ denote the L2 norm of the vector $\boldsymbol{x}$. We assume that

**Assumption 1.** *$F_{t0}$ is continuously differentiable and strictly increasing for all $t \in \mathbb{N}$.*

**Assumption 2.** *$\boldsymbol{b}_{t0} \neq 0$ for all $t \in \mathbb{N}$.*

**Assumption 3.** *The model parameters are normalized such that $F_{t0}(0) = \frac{1}{2}$ for all $t \in \mathbb{N}$, $F_{t0}(1) = p > \frac{1}{2}$ for all $t \in \mathbb{N}$, $\boldsymbol{\theta}_{10} = (1, 0, \ldots, 0)$, $\boldsymbol{\theta}_{20} = (0, 1, 0, \ldots, 0), \ldots, \boldsymbol{\theta}_{D0} = (0, \ldots, 0, 1)$, and $\boldsymbol{\theta}_{D+1,0} = (0, \ldots, 0)$.*

**Assumption 4.** *For any $\varepsilon > 0$ and $\boldsymbol{x} \in \mathbb{R} \times \mathbb{R}_+^D$, there exists a $t \in \mathbb{N}$ such that $\|(a_{t0}, \boldsymbol{b}_{t0}) - \boldsymbol{x}\| < \varepsilon$. For any $\varepsilon > 0$ and $\boldsymbol{x} \in \mathbb{R}^D$, there exists an $n \in \mathbb{N}$ such that $\|\boldsymbol{\theta}_{n0} - \boldsymbol{x}\| < \varepsilon$.*

The first assumption guarantees that the distribution of error terms admits a density with full support. The second assumption states there are no items that fail to distinguish among individuals. The assumptions on $F_{t0}$ and $\boldsymbol{\theta}_{n0}$ constitute normalizations. The normalizations are without loss of generality and are chosen to simplify the proofs. We need to normalize $\boldsymbol{\theta}_{n0}$ in order to prevent translations and rotations of the individual traits. We can equivalently assume that the individual traits $\boldsymbol{\theta}_{n0}$ span $\mathbb{R}^{\mathbb{D}}$ and that the individual traits take on at least $D + 1$ distinct values. The normalizations on $F_{t0}$ are necessary because the location and scale of $F_{t0}$ can already be shifted by the item parameters $a_{t0}$ and $\boldsymbol{b}_{t0}$. We could equivalently assume that $F_{t0}$ has mean zero and variance one, but this alternative normalization would lead to slightly more complicated proofs. The normalizations we choose are simply the most convenient.

The last assumption states that a subsequence of $\{a_{t0}, \boldsymbol{b}_{t0}\}_{t=1}^{\infty}$ converges to $\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R} \times \mathbb{R}_+^D$ and that a subsequence of $\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}$ converges to $\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^D$. In order for this definition to make sense, we need to ensure that there exist such sequences. Recall that the rational numbers are countable (and hence can be represented as a sequence). Furthermore, the rationals are dense in the real line. Thus the rational numbers satisfy this condition. Essentially, this condition requires that the individual characteristics and item parameters "cover" the whole space.

Assumption 4 may initially appear restrictive, especially compared to conditions for the identification of parametric models. We provide three reasons for why this assumption should not be particularly troubling. First, the assumption is not the weakest possible. In online Appendix B (see Online Supplementary Material) we provide a proof of Proposition 1 (see below) under a much weaker assumption to demonstrate that this assumption can be substantially relaxed. The result retains the requirement that $a_{t0}$ cover the real line and $\boldsymbol{b}_{t0}$ cover the positive real line, but allows the individual traits to take on as few as $D + 2$ distinct values. Second, conditions for nonparametric identification will, in general, be more restrictive than conditions for parametric identification. For example, Douglas's (1997) estimator requires that the individual traits have a distribution with full support and requires an infinite number of items. The weaker condition we require in online Appendix B is roughly analogous, after reversing the roles of individuals and items. Third, efficient test design will generally require many different values for the item parameters. Adaptive tests will ensure that the item parameters take on more values than there are items in the test, helping identification further.

We view our results as suggesting that obtaining interval scale estimates of individual traits in a semiparametric framework is possible. While further identification results might be established under weaker conditions, we believe it will be more profitable to implement various estimators and evaluate their properties using Monte Carlo experiments. Identification results alone will never be able to establish how much variation in the individual and item parameters is sufficient for obtaining reliable estimates.

A final potential concern is that our results apply to the fixed effects framework, while the random effects framework is more widely used. Though the random effects framework is currently dominant, the fixed effects framework has been used in the past. Lord's (1980) influential book works in the fixed effects framework. The literature moved away from the fixed effects framework due to the work of Bock and Lieberman (1970) and Bock and Aitken (1981), who demonstrated that one could obtain estimates of individual traits from short tests by combining a random effects framework (which generates consistent estimates of item parameters in short tests with many individuals) with a posteriori estimates of the individual traits based on the estimated item parameters. This will not yield consistent estimates of individuals traits in finite length tests, but will allow for correct inferences to be drawn. This move was made for good reasons, but identification of nonparametric random effects models requires long tests (Douglas, 1997, 2001). Thus, one of the main advantages of the random effects framework for parametric estimation does not apply to the nonparametric case.

### 3.2. *The Homogeneous Case*

In the homogeneous case, the model parameters are given by $(\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}, \{a_{t0}, \boldsymbol{b}_{t0}\}_{t=1}^{\infty}, F_0)$.

**Definition 1.** The parameters of model (9), $(\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}, \{a_{t0}, \boldsymbol{b}_{t0}\}_{t=1}^{\infty}, F_0)$, are *identified* if there does not exist $(\{\boldsymbol{\theta}_{n}\}_{n=1}^{\infty}, \{a_{t}, \boldsymbol{b}_{t}\}_{t=1}^{\infty}, F) \neq (\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}, \{a_{t0}, \boldsymbol{b}_{t0}\}_{t=1}^{\infty}, F_0)$ with $F$ continuously differentiable and strictly increasing, $F(0) = \frac{1}{2}$, $F(1) = p$, $\boldsymbol{\theta}_1 = (1, 0, \ldots, 0)$, $\boldsymbol{\theta}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\boldsymbol{\theta}_D = (0, \ldots, 0, 1)$, and $\boldsymbol{\theta}_{D+1} = (0, \ldots, 0)$, such that

$$F\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n\big) = F_0\big(a_{t0} + \boldsymbol{b}_{t0}'\boldsymbol{\theta}_{n0}\big) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{11}$$

**Proposition 1.** *Suppose that Assumptions 1 through 4 hold. Then the item response model is identified in the homogeneous case.*

*Proof:* Consider any $(\{\boldsymbol{\theta}_{n}\}_{n=1}^{\infty}, \{a_{t}, \boldsymbol{b}_{t}\}_{t=1}^{\infty}, F)$ satisfying the above conditions. Then we must have,

$$a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n = \big(F^{-1} \circ F_0\big)\big(a_{t0} + \boldsymbol{b}_{t0}'\boldsymbol{\theta}_{n0}\big) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{12}$$

Here, $F^{-1}$ must exist since $F$ is strictly increasing. Define

$$m = F^{-1} \circ F_0. \tag{13}$$

Applying (13) to (12), we have

$$a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n = m\big(a_{t0} + \boldsymbol{b}_{t0}'\boldsymbol{\theta}_{n0}\big) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{14}$$

Applying the normalizations on $\boldsymbol{\theta}$ to (14), we have

$$a_t + b_{td} = m(a_{t0} + b_{td0}) \quad \text{for all } d \in \{1, \ldots, D\} \text{ and } t \in \mathbb{N}, \tag{15}$$

$$a_t = m(a_{t0}) \quad \text{for all } t \in \mathbb{N}. \tag{16}$$

Subtracting (16) from (15) for each $d \in \{1, \ldots, D\}$, we have

$$b_{td} = m(a_{t0} + b_{td0}) - m(a_{t0}) \quad \text{for all } d \in \{1, \ldots, D\} \text{ and } t \in \mathbb{N}. \tag{17}$$

We can plug (16) and (17) into (14) to obtain

$$m(a_{t0}) + \sum_{d=1}^{D}\big[m(a_{t0} + b_{td0}) - m(a_{t0})\big]\theta_{nd} - m\big(a_{t0} + \boldsymbol{b}_{t0}'\boldsymbol{\theta}_{n0}\big) = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{18}$$

By assumption, for all $(x, y_1, \ldots, y_D) \in \mathbb{R} \times \mathbb{R}_+^D$ and $\varepsilon > 0$, there exists a $t \in \mathbb{N}$ such that $\|(x, y_1, \ldots, y_D) - (a_{t0}, \boldsymbol{b}_{t0})\| < \varepsilon$. Lemma 1 in Appendix A implies that $m$ is continuous. This implies that the function

$$M(x, y; n) = m(x) + \sum_{d=1}^{D} \big[m(x + y_d) - m(x)\big]\theta_{nd} - m\big(x + \boldsymbol{y}'\boldsymbol{\theta}_{n0}\big) \tag{19}$$

is continuous in $(x, \boldsymbol{y})$ for all $(x, \boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}_+^D$ and $n \in \mathbb{N}$. Lemma 2 in Appendix A implies that

$$m(x) + \sum_{d=1}^{D} \big[m(x + y_d) - m(x)\big]\theta_{nd} = m\big(x + \boldsymbol{y}'\boldsymbol{\theta}_{n0}\big) \quad \text{for all } n \in \mathbb{N} \text{ and } (x, \boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}_+^D. \tag{20}$$

Notice that

$$m(0) = F^{-1}\big(F_0(0)\big) = F^{-1}(1/2) = 0, \tag{21}$$

$$m(1) = F^{-1}\big(F_0(1)\big) = F^{-1}(p) = 1. \tag{22}$$

Define $\boldsymbol{e}_d \in \mathbb{R}^D$ to be the $d$th unit vector. Using $x = 0$ and $\boldsymbol{y} = \boldsymbol{e}_d$ in (20), we have

$$m(0) + \big[m(1) - m(0)\big]\theta_{nd} = m(\theta_{nd0}) \quad \text{for all } n \in \mathbb{N} \text{ and } d \in \{1, \ldots, D\}. \tag{23}$$

Using (21) and (22) in (23), we have

$$\theta_{n,d} = m(\theta_{n,d,0}) \quad \text{for all } n \in \mathbb{N} \text{ and } d \in \{1, \ldots, D\}. \tag{24}$$

Plugging (24) into (18), we obtain

$$m(x) + \sum_{d=1}^{D} \big[m(x + y_d) - m(x)\big]m(\theta_{nd0}) = m\big(x + \boldsymbol{y}'\boldsymbol{\theta}_{n0}\big) \quad \text{for all } n \in \mathbb{N} \text{ and } (x, \boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}_+^D. \tag{25}$$

We can apply Lemma 2 in Appendix A to the sequence $\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}$ to obtain

$$m(x) + \sum_{d=1}^{D} \big[m(x + y_d) - m(x)\big]m(z) = m\big(x + \boldsymbol{y}'\boldsymbol{z}\big) \quad \text{for all } (x, \boldsymbol{y}, \boldsymbol{z}) \in \mathbb{R} \times \mathbb{R}_+^D \times \mathbb{R}^D. \tag{26}$$

Using $z_d = 0$ for $d \in \{2, \ldots, D\}$ in (26), we have

$$m(x) + \big[m(x + y_1) - m(x)\big]m(z_1) = m(x + y_1 z_1) \quad \text{for all } (x, y_1, z_1) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}. \tag{27}$$

Lemma 4 in Appendix A implies that $m(x) = x$, or that $(F^{-1} \circ F_0)(x) = x$. Hence,

$$F_0(x) = F(x) \quad \text{for all } x \in \mathbb{R}. \tag{28}$$

Applying (28) to (11), we have

$$a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n = a_{t0} + \boldsymbol{b}_{t0}'\boldsymbol{\theta}_{n0} \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{29}$$

We use the fact that $\boldsymbol{\theta}_{D+1} = \boldsymbol{\theta}_{D+1,0} = 0$ to determine that $a_t = a_{t0}$ for all $t \in \mathbb{N}$. Then we can use the fact that $\boldsymbol{\theta}_d = \boldsymbol{\theta}_{d0} = \boldsymbol{e}_d$ for $d \in \{1, \ldots, D\}$ to determine that $b_{td} = b_{td0}$ for all $t \in \mathbb{N}$ and $d \in \{1, \ldots, D\}$. This implies that

$$\boldsymbol{b}_{t0}'(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{30}$$

Assumption 4 implies that the vectors $\{\boldsymbol{b}_{t0}\}_{t=1}^{\infty}$ span $\mathbb{R}^D$. Hence, it follows that $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n0}$ for all $n \in \mathbb{N}$, proving the result. □

### 3.3. *The Heterogeneous Case*

Here, we consider the case where the distribution of error terms varies across items.

**Definition 2.** The parameters of model (10), $(\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}, \{a_{t0}, \boldsymbol{b}_{t0}\}_{t=1}^{\infty}, \{F_{t0}\}_{t=1}^{\infty})$, are *identified* if there does not exist $(\{\boldsymbol{\theta}_{n}\}_{n=1}^{\infty}, \{a_{t}, \boldsymbol{b}_{t}\}_{t=1}^{\infty}, \{F_{t}\}_{t=1}^{\infty}) \neq (\{\boldsymbol{\theta}_{n0}\}_{n=1}^{\infty}, \{a_{t0}, \boldsymbol{b}_{t0}\}_{t=1}^{\infty}, \{F_{t0}\}_{t=1}^{\infty})$ with $F_t$ continuously differentiable and strictly increasing, $F_t(0) = \frac{1}{2}$, $F_t(1) = p$, $\boldsymbol{\theta}_1 = (1, 0, \dots, 0)$, $\boldsymbol{\theta}_2 = (0, 1, 0, \dots, 0), \dots, \boldsymbol{\theta}_D = (0, \dots, 0, 1)$, and $\boldsymbol{\theta}_{D+1} = (0, \dots, 0)$, such that

$$F_t\big(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_n\big) = F_{t0}\big(a_{t0} + \boldsymbol{b}_{t0}'\boldsymbol{\theta}_{n0}\big) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{31}$$

We prove that the general unidimensional model is not identified. We show, however, that the individual traits can be identified up to a monotonic transformation.

**Proposition 2.** *Suppose that $D = 1$. Consider any $(\{\theta_n\}_{n=1}^{\infty}, \{a_t, b_t\}_{t=1}^{\infty}, \{F_t\}_{t=1}^{\infty})$ with $F_t$ continuously differentiable and strictly increasing, $F_t(0) = \frac{1}{2}$, $F_t(1) = p$, $\theta_1 = 1$, and $\theta_2 = 0$. Then $(\{\theta_n\}_{n=1}^{\infty}, \{a_t, b_t\}_{t=1}^{\infty}, \{F_t\}_{t=1}^{\infty})$ satisfies (31) if and only if there exists a monotonically increasing transformation $h$ with $h(0) = 0$ and $h(1) = 1$ such that the following hold*:

$$\theta_n = h(\theta_{n0}) \quad \textit{for all } n \in \mathbb{N}, \tag{32}$$

$$a_t = \frac{-h(\frac{-a_{t0}}{b_{t0}})}{h(\frac{1-a_{t0}}{b_{t0}}) - h(\frac{-a_{t0}}{b_{t0}})} \quad \textit{for all } t \in \mathbb{N}, \tag{33}$$

$$b_t = \frac{1}{h(\frac{1-a_{t0}}{b_{t0}}) - h(\frac{-a_{t0}}{b_{t0}})} \quad \textit{for all } t \in \mathbb{N}, \tag{34}$$

$$F_t(y) = F_{t0}\left(a_{t0} + b_{t0}h^{-1}\left(h\left(\frac{-a_{t0}}{b_{t0}}\right) + \left[h\left(\frac{1-a_{t0}}{b_{t0}}\right) - h\left(\frac{-a_{t0}}{b_{t0}}\right)\right]y\right)\right) \quad \textit{for all } t \in \mathbb{N}. \tag{35}$$

*Proof:* Consider any $(\{\theta_n\}_{n=1}^{\infty}, \{a_t, b_t\}_{t=1}^{\infty}, \{F_t\}_{t=1}^{\infty})$ satisfying the above conditions. Then we must have

$$a_t + b_t\theta_n = \big(F_t^{-1} \circ F_{t0}\big)(a_{t0} + b_{t0}\theta_{n0}) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{36}$$

Define

$$m_t = F_t^{-1} \circ F_{t0} \quad \text{for all } t \in \mathbb{N}. \tag{37}$$

Applying (37) to (36), we have

$$a_t + b_t\theta_n = m_t(a_{t0} + b_{t0}\theta_{n0}) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{38}$$

Applying the normalization on $\theta$ to (38), we have

$$a_t + b_t = m_t(a_{t0} + b_{t0}) \quad \text{for all } t \in \mathbb{N}, \tag{39}$$

$$a_t = m_t(a_{t0}) \quad \text{for all } t \in \mathbb{N}. \tag{40}$$

Subtracting (40) from (39), we have

$$b_t = m_t(a_{t0} + b_{t0}) - m_t(a_{t0}) \quad \text{for all } t \in \mathbb{N}. \tag{41}$$

We can plug (40) and (41) into (38) to obtain

$$\big[m_t(a_{t0} + b_{t0}) - m_t(a_{t0})\big]\theta_n = m_t(a_{t0} + b_{t0}\theta_{n0}) - m_t(a_{t0}) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{42}$$

We can rearrange (42) to obtain

$$\theta_n = \frac{m_t(a_{t0} + b_{t0}\theta_{n0}) - m_t(a_{t0})}{m_t(a_{t0} + b_{t0}) - m_t(a_{t0})} \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{43}$$

Define

$$h(x) = \frac{m_1(a_{10} + b_{10}x) - m_1(a_{10})}{m_1(a_{10} + b_{10}) - m_1(a_{10})} \quad \text{for all } x \in \mathbb{R}. \tag{44}$$

Notice that $h(x)$ is monotonically increasing and satisfies $h(0) = 0$ and $h(1) = 1$. Hence, there must exist a monotonically increasing $h$ with $h(0) = 0$ and $h(1) = 1$ such that

$$\theta_n = h(\theta_{n0}) \quad \text{for all } n \in \mathbb{N}. \tag{45}$$

We can plug (40), (41), and (45) into (31) to obtain

$$F_t\Big(m_t(a_{t0}) + \big[m_t(a_{t0} + b_{t0}) - m_t(a_{t0})\big]h(\theta_{n0})\Big) = F_{t0}(a_{t0} + b_{t0}\theta_{n0}) \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathbb{N}. \tag{46}$$

Define

$$M(x;t) = F_t\Big(m_t(a_{t0}) + \big[m_t(a_{t0} + b_{t0}) - m_t(a_{t0})\big]h(x)\Big) - F_{t0}(a_{t0} + b_{t0}x) \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{47}$$

Continuity of $M$ in $x$ for all $t \in \mathbb{N}$ and $x \in \mathbb{R}$ follows from continuity of $F_{t0}$, $F_t$, $m_t$, and $h$. Hence, we can apply Lemma 2 in Appendix A to obtain

$$F_t\Big(m_t(a_{t0}) + \big[m_t(a_{t0} + b_{t0}) - m_t(a_{t0})\big]h(x)\Big) = F_{t0}(a_{t0} + b_{t0}x) \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{48}$$

Using (40) and (41) in (48), we have

$$F_t\big(a_t + b_t h(x)\big) = F_{t0}(a_{t0} + b_{t0}x) \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{49}$$

Using the transformation, $y = a_t + b_t h(x)$ in (49), we have

$$F_t(y) = F_{t0}\left(a_{t0} + b_{t0}h^{-1}\left(\frac{y - a_t}{b_t}\right)\right) \quad \text{for all } t \in \mathbb{N} \text{ and } y \in \mathbb{R}. \tag{50}$$

Notice that

$$\frac{1}{2} = F_t(0) = F_{t0}\left(a_{t0} + b_{t0}h^{-1}\left(\frac{-a_t}{b_t}\right)\right) \quad \text{for all } t \in \mathbb{N}, \tag{51}$$

$$p = F_t(1) = F_{t0}\left(a_{t0} + b_{t0}h^{-1}\left(\frac{1 - a_t}{b_t}\right)\right) \quad \text{for all } t \in \mathbb{N}. \tag{52}$$

Using the fact that $F_{t0}^{-1}(\frac{1}{2}) = 0$ and $F_{t0}^{-1}(p) = 1$, we have

$$0 = a_{t0} + b_{t0}h^{-1}\left(\frac{-a_t}{b_t}\right) \quad \text{for all } t \in \mathbb{N}, \tag{53}$$

$$1 = a_{t0} + b_{t0}h^{-1}\left(\frac{1-a_t}{b_t}\right) \quad \text{for all } t \in \mathbb{N}. \tag{54}$$

We can solve these for $a_t$ and $b_t$ to obtain (33) and (34). We can plug these into (50) to obtain (35). Finally, we can immediately verify that (32), (33), (34), and (35) imply that condition (10) and the normalizations hold simply by plugging these four conditions into the relevant equations. □

Proposition 2 shows that we can only identify the individual characteristics up to a monotonic transformation in the heterogeneous unidimensional case. The result shows that the negative result of Douglas (2001) carries over into the fixed effects framework. Below, we show that we can fully identify the individual characteristics, the item parameters, and the distributions of the error terms, in the multidimensional case.

**Proposition 3.** *Suppose that $D = 2$ and that Assumptions 1 through 4 hold. Then the item response model is identified in the heterogeneous case.*

*Proof:* Define

$$m_t = F_t^{-1} \circ F_{t0} \quad \text{for all } t \in \mathbb{N}. \tag{55}$$

We have

$$a_t + b_{t1}\theta_{n1} + b_{t2}\theta_{n2} = m_t(a_{t0} + b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) \quad \text{for all } (n,t) \in \mathbb{N}^2. \tag{56}$$

Using

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10} = (1, 0), \tag{57}$$

$$\boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20} = (0, 1), \tag{58}$$

$$\boldsymbol{\theta}_3 = \boldsymbol{\theta}_{30} = (0, 0), \tag{59}$$

we have

$$a_t = m_t(a_{t0}) \quad \text{for all } t \in \mathbb{N}, \tag{60}$$

$$a_t + b_{t1} = m_t(a_{t0} + b_{t10}) \quad \text{for all } t \in \mathbb{N}, \tag{61}$$

$$a_t + b_{t2} = m_t(a_{t0} + b_{t20}) \quad \text{for all } t \in \mathbb{N}. \tag{62}$$

Plugging (60) into (61) and (62), we have

$$b_{t1} = m_t(a_{t0} + b_{t10}) - m_t(a_{t0}) \quad \text{for all } t \in \mathbb{N}, \tag{63}$$

$$b_{t2} = m_t(a_{t0} + b_{t20}) - m_t(a_{t0}) \quad \text{for all } t \in \mathbb{N}. \tag{64}$$

We can plug (60), (63), and (64) into (56) to obtain

$$\begin{aligned} &m_t(a_{t0}) + \big[m_t(a_{t0} + b_{t10}) - m_t(a_{t0})\big]\theta_{n1} + \big[m_t(a_{t0} + b_{t20}) - m_t(a_{t0})\big]\theta_{n2} \\ &\quad = m_t(a_{t0} + b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) \quad \text{for all } (n,t) \in \mathbb{N}^2. \end{aligned} \tag{65}$$

Define,

$$\tilde{m}_t(x) = m_t(a_{t0} + x) - m_t(a_{t0}) \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{66}$$

Notice that

$$\tilde{m}_t(0) = 0 \quad \text{for all } t \in \mathbb{N}. \tag{67}$$

We can use (66) in (65) to obtain

$$\tilde{m}_t(b_{t10})\theta_{n1} + \tilde{m}_t(b_{t20})\theta_{n2} = \tilde{m}_t(b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) \quad \text{for all } (n,t) \in \mathbb{N}^2. \tag{68}$$

We can stack (68) for two different values of $t$ to obtain

$$\begin{bmatrix} \tilde{m}_s(b_{s10}) & \tilde{m}_s(b_{s20}) \\ \tilde{m}_t(b_{t10}) & \tilde{m}_t(b_{t20}) \end{bmatrix} \begin{bmatrix} \theta_{n1} \\ \theta_{n2} \end{bmatrix} = \begin{bmatrix} \tilde{m}_s(b_{s10}\theta_{n10} + b_{s20}\theta_{n20}) \\ \tilde{m}_t(b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) \end{bmatrix}$$
$$\text{for all } (n,s,t) \in \mathbb{N}^3. \tag{69}$$

Let $\mathcal{T}^*$ denote the set of $(s,t,u) \in \mathbb{N}^3$ with $s < t < u$ satisfying

$$b_{s10} > b_{s20} > 0, \tag{70}$$

$$b_{t20} > b_{t10} > 0, \tag{71}$$

$$b_{u20} > 0 > b_{u10}. \tag{72}$$

Since $F_{s0}$ and $F_{t0}$ are strictly increasing, (70) and (71) imply

$$F_{s0}(b_{s10}) > F_{s0}(b_{s20}) > F_{s0}(0) = \frac{1}{2} \quad \text{for all } (s,t,u) \in \mathcal{T}^*, \tag{73}$$

$$F_{t0}(b_{t20}) > F_{t0}(b_{t10}) > F_{t0}(0) = \frac{1}{2} \quad \text{for all } (s,t,u) \in \mathcal{T}^*. \tag{74}$$

Since $F_s^{-1}$ and $F_t^{-1}$ are strictly increasing as well, we have

$$F_s^{-1}\big(F_{s0}(b_{s10})\big) > F_s^{-1}\big(F_{s0}(b_{s20})\big) > F_s^{-1}\big(F_{s0}(0)\big) = 0 \quad \text{for all } (s,t,u) \in \mathcal{T}^*, \tag{75}$$

$$F_t^{-1}\big(F_{t0}(b_{t20})\big) > F_t^{-1}\big(F_{t0}(b_{t10})\big) > F_t^{-1}\big(F_{t0}(0)\big) = 0 \quad \text{for all } (s,t,u) \in \mathcal{T}^*, \tag{76}$$

which implies that

$$\tilde{m}_s(b_{s10}) > \tilde{m}_s(b_{s20}) > 0 \quad \text{for all } (s,t,u) \in \mathcal{T}^*, \tag{77}$$

$$\tilde{m}_t(b_{t20}) > \tilde{m}_t(b_{t10}) > 0 \quad \text{for all } (s,t,u) \in \mathcal{T}^*. \tag{78}$$

Next,

$$\begin{aligned} &\tilde{m}_s(b_{s10})\tilde{m}_t(b_{t20}) - \tilde{m}_s(b_{s20})\tilde{m}_t(b_{t10}) \\ &= \big(\tilde{m}_s(b_{s10}) - \tilde{m}_s(b_{s20})\big)\tilde{m}_t(b_{t20}) + \tilde{m}_s(b_{s20})\big(\tilde{m}_t(b_{t20}) - \tilde{m}_t(b_{t10})\big) > 0 \\ &\text{for all } (s,t,u) \in \mathcal{T}^*. \end{aligned} \tag{79}$$

Similarly, we can use (70), (71), and (72) to determine that

$$\tilde{m}_u(b_{u10})\tilde{m}_t(b_{t20}) - \tilde{m}_u(b_{u20})\tilde{m}_t(b_{t10}) < 0 \quad \text{for all } (s,t,u) \in \mathcal{T}^*, \tag{80}$$

$$\tilde{m}_u(b_{u20})\tilde{m}_s(b_{s10}) - \tilde{m}_u(b_{u10})\tilde{m}_s(b_{s20}) > 0 \quad \text{for all } (s,t,u) \in \mathcal{T}^*. \tag{81}$$

Now, (79) implies that the matrix in (69) is invertible. Hence, we have

$$\begin{bmatrix}\theta_{n1}\\ \theta_{n2}\end{bmatrix} = \frac{1}{\tilde{m}_s(b_{s10})\tilde{m}_t(b_{t20}) - \tilde{m}_s(b_{s20})\tilde{m}_t(b_{t10})} \times \begin{bmatrix}\tilde{m}_t(b_{t20}) & -\tilde{m}_s(b_{s20})\\ -\tilde{m}_t(b_{t10}) & \tilde{m}_s(b_{s10})\end{bmatrix}\begin{bmatrix}\tilde{m}_s(b_{s10}\theta_{n10} + b_{s20}\theta_{n20})\\ \tilde{m}_t(b_{t10}\theta_{n10} + b_{t20}\theta_{n20})\end{bmatrix}$$
$$\text{for all } n \in \mathbb{N} \quad \text{for all } (s,t,u) \in \mathcal{T}^*. \tag{82}$$

We can plug this into (68) to obtain

$$\frac{\tilde{m}_u(b_{u10})\tilde{m}_t(b_{t20}) - \tilde{m}_u(b_{u20})\tilde{m}_t(b_{t10})}{\tilde{m}_s(b_{s10})\tilde{m}_t(b_{t20}) - \tilde{m}_s(b_{s20})\tilde{m}_t(b_{t10})}\tilde{m}_s(b_{s10}\theta_{n10} + b_{s20}\theta_{n20}) + \frac{\tilde{m}_u(b_{u20})\tilde{m}_s(b_{s10}) - \tilde{m}_u(b_{u10})\tilde{m}_s(b_{s20})}{\tilde{m}_s(b_{s10})\tilde{m}_t(b_{t20}) - \tilde{m}_s(b_{s20})\tilde{m}_t(b_{t10})}\tilde{m}_t(b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) = \tilde{m}_u(b_{u10}\theta_{n10} + b_{u20}\theta_{n20}) \quad \text{for all } n \in \mathbb{N} \text{ and } (s,t,u) \in \mathcal{T}^*. \tag{83}$$

For simplicity, we define the constants

$$c_{stu1} = \frac{\tilde{m}_u(b_{u10})\tilde{m}_t(b_{t20}) - \tilde{m}_u(b_{u20})\tilde{m}_t(b_{t10})}{\tilde{m}_s(b_{s10})\tilde{m}_t(b_{t20}) - \tilde{m}_s(b_{s20})\tilde{m}_t(b_{t10})} \quad \text{for all } (s,t,u) \in \mathcal{T}^*, \tag{84}$$

$$c_{stu2} = \frac{\tilde{m}_u(b_{u20})\tilde{m}_s(b_{s10}) - \tilde{m}_u(b_{u10})\tilde{m}_s(b_{s20})}{\tilde{m}_s(b_{s10})\tilde{m}_t(b_{t20}) - \tilde{m}_s(b_{s20})\tilde{m}_t(b_{t10})} \quad \text{for all } (s,t,u) \in \mathcal{T}^*. \tag{85}$$

This gives

$$c_{stu1}\tilde{m}_s(b_{s10}\theta_{n10} + b_{s20}\theta_{n20}) + c_{stu2}\tilde{m}_t(b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) - \tilde{m}_u(b_{u10}\theta_{n10} + b_{u20}\theta_{n20}) = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } (s,t,u) \in \mathcal{T}^*. \tag{86}$$

Now, consider the function

$$M_\theta(x,y;n,s,t,u) = c_{stu1}\tilde{m}_s(b_{s10}x + b_{s20}y) + c_{stu2}\tilde{m}_t(b_{t10}x + b_{t20}y) - \tilde{m}_u(b_{u10}x + b_{u20}y) = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } (s,t,u) \in \mathcal{T}^*. \tag{87}$$

Now, $M_\theta$ is continuous in $(x, y)$ for all $n \in \mathbb{N}$, $(s,t,u) \in \mathcal{T}^*$, and $(x, y) \in \mathbb{R}^2$. Hence, Lemma 2 in Appendix A implies that

$$c_{stu1}\tilde{m}_s(b_{s10}x + b_{s20}y) + c_{stu2}\tilde{m}_t(b_{t10}x + b_{t20}y) = \tilde{m}_u(b_{u10}x + b_{u20}y) \quad \text{for all } n \in \mathbb{N}, (s,t,u) \in \mathcal{T}^*, \text{ and } (x,y) \in \mathbb{R}^2. \tag{88}$$

Let us select $y = -\frac{b_{s10}}{b_{s20}}x$. We can plug this into (88) to obtain

$$c_{stu2}\tilde{m}_t\left(\left[b_{t10} - b_{t20}\frac{b_{s10}}{b_{s20}}\right]x\right) = \tilde{m}_u\left(\left[b_{u10} - b_{u20}\frac{b_{s10}}{b_{s20}}\right]x\right) \quad \text{for all } n \in \mathbb{N}, (s,t,u) \in \mathcal{T}^*, \text{ and } x \in \mathbb{R}. \tag{89}$$

Plugging $z = [b_{t10} - \frac{b_{s10}}{b_{s20}}b_{t20}]x$ into (89) yields

$$\tilde{m}_t(z) = \frac{1}{c_{stu2}}\tilde{m}_u\left(\frac{b_{s20}b_{u10} - b_{s10}b_{u20}}{b_{s20}b_{t10} - b_{s10}b_{t20}}z\right) \quad \text{for all } (s,t,u) \in \mathcal{T}^* \text{ and } z \in \mathbb{R}. \tag{90}$$

Similarly, we can select $y = -\frac{b_{t10}}{b_{t20}}x$ and $z = [b_{s10} - \frac{b_{t10}}{b_{t20}}b_{s20}]x$ and obtain

$$\tilde{m}_s(z) = \frac{1}{c_{stu1}}\tilde{m}_t\left(\frac{b_{t20}b_{u10} - b_{t10}b_{u20}}{b_{t20}b_{s10} - b_{t10}b_{s20}}z\right) \quad \text{for all } (s,t,u) \in \mathcal{T}^* \text{ and } z \in \mathbb{R}. \tag{91}$$

We can plug (90) and (91) into (88) to obtain

$$\begin{aligned}
&\tilde{m}_u\left(\frac{b_{u10}b_{t20} - b_{t10}b_{u20}}{b_{s10}b_{t20} - b_{t10}b_{s20}}[b_{s10}x + b_{s20}y]\right) \\
&\quad + \tilde{m}_u\left(\frac{b_{u10}b_{s20} - b_{s10}b_{u20}}{b_{t10}b_{s20} - b_{s10}b_{t20}}[b_{t10}x + b_{t20}y]\right) = \tilde{m}_u(b_{u10}x + b_{u20}y) \\
&\quad \text{for all } (s,t,u) \in \mathcal{T}^* \text{ and } (x,y) \in \mathbb{R}^2.
\end{aligned} \tag{92}$$

Which further implies that

$$\begin{aligned}
&m_u\left(\frac{b_{u10}b_{s20} - b_{s10}b_{u20}}{b_{t10}b_{s20} - b_{s10}b_{t20}}[b_{t10}x + b_{t20}y]\right) \\
&\quad + m_u\left(\frac{b_{u10}b_{t20} - b_{t10}b_{u20}}{b_{s10}b_{t20} - b_{t10}b_{s20}}[b_{s10}x + b_{s20}y]\right) \\
&= m_u(a_{u0}) + m_u(b_{u10}x + b_{u20}y) \quad \text{for all } (s,t,u) \in \mathcal{T}^* \text{ and } (x,y) \in \mathbb{R}^2.
\end{aligned} \tag{93}$$

Define

$$A_\delta = \left\{(\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}) : \delta_{11} > \delta_{12} > 0, \delta_{22} > \delta_{21} > 0\right\}, \tag{94}$$

$$B_\delta = \left\{(\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}) : \delta_{12} = \delta_{21} = 0, \delta_{11} > 0, \delta_{22} > 0\right\}, \tag{95}$$

$$\begin{aligned}
&M_\delta(\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}; s, t, u, x, y) \\
&= m_u\left(\frac{b_{u10}\delta_{22} - \delta_{21}b_{u20}}{\delta_{11}\delta_{22} - \delta_{21}\delta_{12}}[\delta_{11}x + \delta_{12}y]\right) + m_u\left(\frac{b_{u10}\delta_{12} - \delta_{11}b_{u20}}{\delta_{21}\delta_{12} - \delta_{11}\delta_{22}}[\delta_{21}x + \delta_{22}y]\right) \\
&\quad - m_u(a_{u0}) + m_u(b_{u10}x + b_{u20}y) \\
&\text{for all } (\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}) \in A_\delta, (s,t,u) \in \mathcal{T}^*, \text{ and } (x,y) \in \mathbb{R}^2.
\end{aligned} \tag{96}$$

Define $\mathcal{U}^* = \{u \in \mathbb{N} : \exists (s,t,u) \in \mathcal{T}^*\}$. Applying Lemma 2 in Appendix A, we have

$$\begin{aligned}
&\tilde{m}_u\left(\frac{b_{u10}\delta_{22} - \delta_{21}b_{u20}}{\delta_{11}\delta_{22} - \delta_{21}\delta_{12}}[\delta_{11}x + \delta_{12}y]\right) \\
&\quad + \tilde{m}_u\left(\frac{b_{u10}\delta_{12} - \delta_{11}b_{u20}}{\delta_{21}\delta_{12} - \delta_{11}\delta_{22}}[\delta_{21}x + \delta_{22}y]\right) = \tilde{m}_u(b_{u10}x + b_{u20}y) \\
&\quad \text{for all } (\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}) \in A_\delta, u \in \mathcal{U}^*, \text{ and } (x,y) \in \mathbb{R}^2.
\end{aligned} \tag{97}$$

Notice that $M_\delta$ satisfies the conditions of Lemma 5 in Appendix A, with $A = A_\delta$ and $B = B_\delta$. Hence, we can apply Lemma 5 in Appendix A to obtain that (97) holds for $(\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}) \in B_\delta$ as well. Selecting $\delta_{12} = \delta_{21} = 0$, we have

$$m_u(a_{u0} + b_{u20}y) + m_u(a_{u0} + b_{u10}x) = m_u(a_{u0}) + m_u(a_{u0} + b_{u10}x + b_{u20}y) \tag{98}$$

$$\text{for all } u \in \mathcal{U}^* \text{ and } (x,y) \in \mathbb{R}^2, \tag{99}$$

or equivalently,

$$\tilde{m}_u(b_{u20}y) + \tilde{m}_u(b_{u10}x) = \tilde{m}_u(b_{u10}x + b_{u20}y) \quad \text{for all } u \in \mathcal{U}^* \text{ and } (x, y) \in \mathbb{R}^2. \tag{100}$$

Using the transformation,

$$y' = b_{u20}y, \tag{101}$$

$$x' = b_{u10}x, \tag{102}$$

we obtain

$$\tilde{m}_u\left(x' + y'\right) = \tilde{m}_u\left(y'\right) + \tilde{m}_u\left(x'\right) \quad \text{for all } u \in \mathcal{U}^* \text{ and } \left(x', y'\right) \in \mathbb{R}^2. \tag{103}$$

Using (103) and the fact that $\tilde{m}_u$ is continuous, Lemma 3 in Appendix A implies that

$$\tilde{m}_u(z) = z \quad \text{for all } z \in \mathbb{R} \text{ and } u \in \mathcal{U}^*. \tag{104}$$

Using (66), we have

$$m_u(z + a_{u0}) - m_u(a_{u0}) = z \quad \text{for all } z \in \mathbb{R} \text{ and } u \in \mathcal{U}^*. \tag{105}$$

Using the transformation, $z' = z + a_{u0}$, we have

$$m_u\left(z'\right) = m_u(a_{u0}) + \left(z' - a_{u0}\right) \quad \text{for all } z' \in \mathbb{R} \text{ and } u \in \mathcal{U}^*. \tag{106}$$

Plugging $z' = a_{u0}$ into (105), we have

$$m_u(a_{u0}) = a_{u0} \quad \text{for all } u \in \mathcal{U}^*. \tag{107}$$

Plugging (107) into (106), we have

$$m_u(z) = z \quad \text{for all } z \in \mathbb{R} \text{ and } u \in \mathcal{U}^*. \tag{108}$$

Using (108) and (55), we have

$$F_{u0}(x) = F_u(x) \quad \text{for all } u \in \mathcal{U}^* \text{ and } x \in \mathbb{R}, \tag{109}$$

or, equivalently,

$$F_{t0}(x) = F_t(x) \quad \text{for all } t \in \mathcal{U}^* \text{ and } x \in \mathbb{R}. \tag{110}$$

Using (60), (63), (64), and (110), we have

$$a_t = a_{t0} \quad \text{for all } t \in \mathcal{U}^*, \tag{111}$$

$$b_{t1} = b_{t10} \quad \text{for all } t \in \mathcal{U}^*, \tag{112}$$

$$b_{t2} = b_{t20} \quad \text{for all } t \in \mathcal{U}^*. \tag{113}$$

Plugging this into (56), we have

$$b_{t10}(\theta_{n1} - \theta_{n10}) + b_{t20}(\theta_{n2} - \theta_{n20}) = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } t \in \mathcal{U}^*. \tag{114}$$

Since the vectors $\{b_{t10}, b_{t20}\}_{t\in\mathcal{U}^*}$ span $\mathbb{R}^2$, we have $(\theta_{n1}, \theta_{n2}) = (\theta_{n10}, \theta_{n20})$ for all $n \in \mathbb{N}$ proving that $\{\theta_{n0}\}_{n=0}^{\infty}$ is identified.

To complete the proof, we must demonstrate that $(a_{t0}, \boldsymbol{b}_{t0}, F_{t0})$ are identified for all $t \in \mathbb{N}$ (rather than simply all $t \in \mathcal{U}^*$). We assume that $b_{t10} \neq 0$ (the case where $b_{t20} \neq 0$ can be proved symmetrically). We can use (56) to determine that

$$F_t(a_t + b_{t1}\theta_{n10} + b_{t2}\theta_{n20}) = F_{t0}(a_{t0} + b_{t10}\theta_{n10} + b_{t20}\theta_{n20}) \quad \text{for all } (n, t) \in \mathbb{N}^2. \tag{115}$$

Using Lemma 1 in Appendix A, we have

$$F_t(a_t + b_{t1}x + b_{t2}y) = F_{t0}(a_{t0} + b_{t10}x + b_{t20}y) \quad \text{for all } t \in \mathbb{N} \text{ and } (x, y) \in R^2. \tag{116}$$

Using $y = 0$, we obtain

$$F_t(a_t + b_{t1}x) = F_{t0}(a_{t0} + b_{t10}x) \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}, \tag{117}$$

or

$$a_t + b_{t1}x = m_t(a_{t0} + b_{t10}x) \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{118}$$

Using $y = a_{t0} + b_{t10}x$, we have

$$m_t(y) = a_t + b_{t1}\frac{y - a_{t0}}{b_{t10}} \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}, \tag{119}$$

which implies that $m_t$ is linear. Using the fact that $m_t(0) = 0$ and $m_t(1) = 1$, we can determine that

$$m_t(x) = x \quad \text{for all } t \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{120}$$

This immediately implies that $F_t(x) = F_{t0}(x)$ for all $t \in \mathbb{N}$ and $x \in \mathbb{R}$. Applying (120) to (60), (63), and (64), we obtain $a_t = a_{t0}$, $b_{t1} = b_{t10}$, and $b_{t2} = b_{t20}$, for all $t \in \mathbb{N}$, proving the result. □

It is reasonable to believe that this result would extend to the $D$-dimensional case, where $D \geq 3$, and that the same proof strategy would carry over. The proof, however, would be significantly longer due to the proliferation of notation. In particular, it would be difficult to derive conditions that guarantee that the $D$-dimensional generalization of the matrix in (69) is invertible.

## 4. Estimation

The results of the previous section suggest that we could develop two new estimators for the item response model. Proposition 1 suggests that we can estimate

$$(\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{N0}, a_{10}, \boldsymbol{b}_{10}, \ldots, a_{T0}, \boldsymbol{b}_{T0}, F_0), \tag{121}$$

where $F_0$ is a common error distribution for all individuals and items. Proposition 3 suggests that in the multidimensional case we can estimate

$$(\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{N0}, a_{10}, \boldsymbol{b}_{10}, \ldots, a_{T0}, \boldsymbol{b}_{T0}, F_{10}, \ldots, F_{T0}) \tag{122}$$

where $F_{t0}$ for $t \in \{1, \ldots, T\}$ is the distribution of the error term common over individuals, but specific to item $t$.

Consider the first case. We could estimate $(\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{N0}, a_{10}, \boldsymbol{b}_{10}, \ldots, a_{T0}, \boldsymbol{b}_{T0}, F)$ by maximizing the likelihood function,

$$L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, F) = \prod_{n=1}^{N} \prod_{t=1}^{T} F\left(a_t + \boldsymbol{b}_t' \boldsymbol{\theta}_n\right)^{y_{n,t}} \left[1 - F\left(a_t + \boldsymbol{b}_t' \boldsymbol{\theta}_n\right)\right]^{1-y_{n,t}}, \tag{123}$$

over the space $\mathbb{R}^{ND+T(D+1)} \times \mathcal{F}$. In principle, $\mathcal{F}$ could be taken to be an infinite dimensional space (see, e.g., Kiefer & Wolfowitz, 1956). Practical considerations, however, suggest employing a relatively large finite dimensional space. Specifically, we consider the parametric family $F(x) = F(x; \boldsymbol{\gamma})$, where $\boldsymbol{\gamma} \in \Gamma \subset \mathbb{R}^J$. In this case, our likelihood function would be

$$L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma}) = \prod_{n=1}^{N} \prod_{t=1}^{T} F\left(a_t + \boldsymbol{b}_t' \boldsymbol{\theta}_n; \boldsymbol{\gamma}\right)^{y_{n,t}} \left[1 - F\left(a_t + \boldsymbol{b}_t' \boldsymbol{\theta}_n; \boldsymbol{\gamma}\right)\right]^{1-y_{n,t}}. \tag{124}$$

For example, we could allow this parametric family to represent a linear spline over the grid of points, $\{\tilde{x}_1, \ldots, \tilde{x}_J\}$ where $\tilde{x}_1 < \tilde{x}_2 < \cdots < \tilde{x}_J$. In this case, we have

$$F(x; \boldsymbol{\gamma}) = \begin{cases} 0, & x \le \tilde{x}_1, \\ \gamma_1 + (\gamma_2 - \gamma_1) \frac{x - \tilde{x}_1}{\tilde{x}_2 - \tilde{x}_1}, & \tilde{x}_1 \le x \le \tilde{x}_2, \\ \cdots & \\ \gamma_{J-1} + (\gamma_J - \gamma_{J-1}) \frac{x - \tilde{x}_{J-1}}{\tilde{x}_J - \tilde{x}_{J-1}}, & \tilde{x}_{J-1} \le x \le \tilde{x}_J, \\ 1, & x \ge \tilde{x}_J. \end{cases} \tag{125}$$

We would have $\Gamma = \{\boldsymbol{\gamma} \in \mathbb{R}^J : \gamma_{j-1} \le \gamma_j - \delta, \gamma_1 = 0, \gamma_J = 1\}$, where $\delta > 0$ is a constant that ensures that the family of distributions has full support. If the grid points are both closely spaced and cover a large portion of the real line, then this parametric space will provide an approximation to the nonparametric space. The choice of linear splines is attractive here because it is relatively easy to impose the constraints necessary to ensure that the class $F(x) = F(x; \boldsymbol{\gamma})$, where $\boldsymbol{\gamma} \in \Gamma$ contains only continuous cumulative distribution functions.

We can define the estimator as

$$\begin{aligned} &(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_N, \hat{a}_1, \hat{\boldsymbol{b}}_1, \ldots, \hat{a}_1, \hat{\boldsymbol{b}}_T, \hat{\boldsymbol{\gamma}}) \\ &\quad = \underset{(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma})}{\arg\max} L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma}). \end{aligned} \tag{126}$$

The likelihood function $L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma})$ must be optimized over a very large parameter space, but we can apply the Zig-Zag algorithm to reduce the computational burden (Heckman, 1981; Poole & Rosenthal, 1997). Specifically, notice that we can alternatively define the estimator as

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \Gamma}{\arg\max} \prod_{n=1}^{N} \prod_{t=1}^{T} F\left(\hat{a}_t + \hat{\boldsymbol{b}}_t' \hat{\boldsymbol{\theta}}_n; \boldsymbol{\gamma}\right)^{y_{n,t}} \left[1 - F\left(\hat{a}_t + \hat{\boldsymbol{b}}_t' \hat{\boldsymbol{\theta}}_n; \boldsymbol{\gamma}\right)\right]^{1-y_{n,t}}, \tag{127}$$

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}_n}{\arg\max} \prod_{t=1}^{T} F\left(\hat{a}_t + \hat{\boldsymbol{b}}_t' \boldsymbol{\theta}_n; \hat{\boldsymbol{\gamma}}\right)^{y_{n,t}} \left[1 - F\left(\hat{a}_t + \hat{\boldsymbol{b}}_t' \boldsymbol{\theta}_n; \hat{\boldsymbol{\gamma}}\right)\right]^{1-y_{n,t}} \quad \text{for } n \in \{1, \ldots, N\}, \tag{128}$$

$$\left(\hat{a}_t, \hat{\boldsymbol{b}}_t\right) = \underset{(a_t, \boldsymbol{b}_t)}{\arg\max} \prod_{n=1}^{N} F\left(a_t + \boldsymbol{b}_t' \hat{\boldsymbol{\theta}}_n; \hat{\boldsymbol{\gamma}}\right)^{y_{n,t}} \left[1 - F\left(a_t + \boldsymbol{b}_t' \hat{\boldsymbol{\theta}}_n; \hat{\boldsymbol{\gamma}}\right)\right]^{1-y_{n,t}} \quad \text{for } t \in \{1, \ldots, T\}. \tag{129}$$

This suggests that we can compute the estimator by repeating the following three steps:

1. Maximize $\boldsymbol{\gamma}$ over $\Gamma$. Note that in the case of the linear spline, this involves constrained optimization because of the requirement that the $\boldsymbol{\gamma}$'s be increasing.
2. For each $n$, one at a time, maximize the portion of the likelihood that depends on $\boldsymbol{\theta}_n$.
3. For each $t$, one at a time, maximize the portion of the likelihood that depends on $(a_t, \boldsymbol{b}_t)$.

The algorithm can terminated at the point at which subsequent parameter values do not change much between iterations. A naive algorithm that employed a quasi-Newton method to maximize the likelihood and employed finite difference methods to compute the gradient would require $O(\max\{N^2T, NT^2\})$ operations per iteration. Using the Zig-Zag algorithm, we instead require $O(NT)$ operations per iteration, making the algorithm more practical.

There is one more practical problem that must we deal with. There is a possibility of observing perfect separation in some of the items or for some of the individuals (e.g., items that everyone answers correctly or individuals who answer all items correctly). In this case, the maximum likelihood estimates will not be well defined. For this reason, we consider maximizing the penalized log-likelihood (Peress & Spirling, 2010),

$$\begin{aligned}&\tilde{l}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma}) \\ &\quad = l(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma}) + \lambda_\theta \sum_{n=1}^{N} \boldsymbol{\theta}_n' \boldsymbol{\theta}_n + \lambda_{ab} \sum_{t=1}^{T} (a_t, \boldsymbol{b}_t)'(a_t, \boldsymbol{b}_t).\end{aligned} \tag{130}$$

Here, $l(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma}) = \log L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma})$ denotes the log-likelihood and $\lambda_\theta > 0$ and $\lambda_{ab} > 0$ are penalty terms. Notice that this estimator corresponds to maximizing a posterior distribution when independent normal priors and placed on $\boldsymbol{\theta}_n$ and $(a_t, \boldsymbol{b}_t)$. Specifically, $\boldsymbol{\theta}_n \sim N(0, (2\lambda_\alpha)^{-1} I_D)$ and $(a_t, \boldsymbol{b}_t) \sim N(0, (2\lambda_{ab})^{-1} I_{D+1})$, where $I_k$ denotes an identity matrix with dimension $k$.

The second estimator suggested by this paper is conceptually similar. Instead of estimating a common $F$ for all items, we estimate a separate $F_t$ for each item. We assume that we have $F_t(x) = F(x; \boldsymbol{\gamma}_t)$ for $t \in \{1, \ldots, T\}$. We can write

$$L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, a_1, \boldsymbol{b}_1, \ldots, a_T, \boldsymbol{b}_T, \boldsymbol{\gamma}) = \prod_{n=1}^{N} \prod_{t=1}^{T} F\left(a_t + \boldsymbol{b}_t' \boldsymbol{\theta}_n; \boldsymbol{\gamma}_t\right)^{y_{n,t}} \left[1 - F\left(a_t + \boldsymbol{b}_t' \boldsymbol{\theta}_n; \boldsymbol{\gamma}_t\right)\right]^{1-y_{n,t}}, \tag{131}$$

where $D > 1$ is required for identification of $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N)$. In this case, we can once again apply the Zig-Zag algorithm to estimate this model, but since $\boldsymbol{\gamma}_t$ are now item-specific rather than common parameters, the Zig-Zag algorithm only has two steps. As before, we recommend that penalized maximum likelihood be applied to deal with the perfect separation that may occur for a small number of items or individuals.

## 5. Conclusions

In this paper, we considered the identification of a semiparametric item response model. In particular, we addressed the concern that interval scale estimates generated by parametric estimators are the result of arbitrary functional form. This concern comes from two sources.

First, existing results suggest that the shape of the distribution of individual characteristics is not identified in the random effects model (Douglas, 2001). Second, the one existing nonparametric fixed effects estimator is only able to produce a rank ordering of individual characteristics in one dimension (Poole, 2000). Our results show that in a semiparametric fixed effects framework, individual and item characteristics are identified under reasonable assumptions.

We suggest two possible consequences of our results. First, our results provide a justification for the continued use of interval scale estimates obtained from parametric estimators. The fact that, previously, interval scale estimates of individual traits were only known to be identified in parametric models was potentially troubling. It suggested the possibility that interval scale estimates were an artifact of assumed functional form. Instead, our results suggest that interval scale estimates can be recovered under much weaker assumptions, such as homogeneity of the error term or a linear compensatory framework. In other words, it is reasonable to "believe" the shape of individual characteristics recovered from parametric models.

In addition to providing additional confidence in the use of parametric estimators, our results suggest that two new types of estimator could be developed. The first estimator would correspond to the homogeneous case (exploiting Proposition 1). In contrast to existing nonparametric estimators (Ramsay & Abrahamowicz, 1989; Ramsay, 1991; Douglas, 1997; Sijtsma, 1998), it would require estimating only one unknown function (i.e., $F_0$) as opposed to one unknown function per item. The second estimator would correspond to a linear compensatory model with heterogeneous errors (exploiting Proposition 3). This estimator would apply to the multidimensional case, for which few nonparametric estimators exist. The estimator would require estimating an unknown function for each item, but would not require estimating multidimensional unknown functions. Exactly how these estimators should be implemented (kernels, sieves, smoothing splines, Dirichlet priors, etc.) is a question for future research, but we suggested one possible approach in Section 4.

It is an open question whether the results in Propositions 1 and 3 would carry over to a random effects framework. Identification would certainly not hold for a finite length test (because we cannot recover an infinite dimensional quantity from a finite number of moments), but it is possible that one could recover the relevant quantities in a long test. In the event that the random effects framework is indeed identified, the hope is that our results will provide some progress towards discovering a proof. Alternatively, if the random effects framework is not identified, even asymptotically, then using a fixed effects estimator would be the cost one would have to pay if one wanted to recover interval scale estimates of individual traits based on homogeneous errors or a multidimensional linear compensatory framework.

## Acknowledgements

I would like to thank Kosuke Imai, William Lowe, Alexander Robitzsch, two anonymous referees, the associate editor, and participants of a panel at the Midwest Political Science Association meetings (Chicago, 2011) for helpful comments and suggestions.

## Appendix A. Proof of Lemmas

**Lemma 1.** *Suppose that $F_0 : \mathbb{R} \to [0,1]$ and $F : \mathbb{R} \to [0,1]$ are continuously differentiable and strictly increasing. Define $m = F^{-1} \circ F_0$. Then $m(x)$ is continuously differentiable for all $x \in \mathbb{R}$.*

*Proof:* Since $F$ is continuously differentiable and strictly increasing, the inverse function theorem implies that $F^{-1}(x)$ exists and is continuous and differentiable for all $x \in (0,1)$. Since the composition of continuous functions is continuous, continuity of $m$ follows from continuity of $F_0$. We can use the formulas for the derivative of an inverse and the derivative of a composition to determine that $m'(x) = \frac{f_0(F^{-1}(x))}{f(F^{-1}(x))}$. Since $f$ and $f_0$ are continuous and $f(y) > 0$ for all $y \in (0, 1)$, we find that $m'$ is continuous, proving the result. $\square$

**Lemma 2.** *Suppose that $f : A \to \mathbb{R}^K$ is continuous where $A \subset \mathbb{R}^J$. Suppose that for all $i \in \mathbb{N}$, $f(\boldsymbol{\theta}_i) = 0$. Suppose that for all $\boldsymbol{x} \in A$ and $\varepsilon > 0$, there exists an $i \in \mathbb{N}$ such that $\|\boldsymbol{x} - \boldsymbol{\theta}_i\| < \varepsilon$. Then for all $\boldsymbol{x} \in A$, $f(\boldsymbol{x}) = 0$.*

*Proof:* If $f$ is continuous, then for each $\boldsymbol{x} \in A$ and $\lambda > 0$, there exists an $\varepsilon > 0$ such that $\|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon$ implies that $\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| < \lambda$. For each $\lambda > 0$, we can select $i \in \mathbb{N}$ such that $\|\boldsymbol{x} - \boldsymbol{\theta}_i\| < \varepsilon$, implying that $\|f(\boldsymbol{x}) - f(\boldsymbol{\theta}_i)\| < \lambda$. By assumption, $f(\boldsymbol{\theta}_i) = 0$, so that $\|f(\boldsymbol{x}) - f(\boldsymbol{\theta}_i)\| = \|f(\boldsymbol{x})\| < \lambda$. Since $\|f(\boldsymbol{x})\| < \lambda$ for all $\lambda > 0$, it must be the case that $f(\boldsymbol{x}) = 0$, proving the result. $\square$

**Lemma 3.** *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous and*

$$f(x + y) = f(x) + f(y) \quad \textit{for } (x, y) \in \mathbb{R} \times \mathbb{R}_+. \tag{A.1}$$

*Then $f(x) = \kappa x$ for all $x \in \mathbb{R}$, where $\kappa = f(1)$.*

*Proof:* Using $x = 0$ and $y = 0$ in (A.1), we have $f(0) = 2f(0)$ implying that $f(0) = 0$. Next, for $x \geq 0$, notice that $0 = f(0) = f((-x) + x) = f(-x) + f(x)$. Hence, we have

$$f(-x) = -f(x) \quad \text{for all } x \in \mathbb{R}_+. \tag{A.2}$$

For all $x \leq 0$, $0 = f(0) = f(x + (-x)) = f(-x) + f(x)$ so that

$$f(-x) = -f(x) \quad \text{for all } x \in \mathbb{R}. \tag{A.3}$$

Consider any positive integers $j$ and $k$. Applying (A.1) recursively, we have

$$f\left(\frac{j}{k}x\right) = f\left(\frac{x}{k} + \frac{x}{k} + \cdots + \frac{x}{k}\right) = jf\left(\frac{x}{k}\right) \quad \text{for all } x \in \mathbb{R}_+. \tag{A.4}$$

We further have

$$kf\left(\frac{x}{k}\right) = f\left(\frac{x}{k}\right) + f\left(\frac{x}{k}\right) + \cdots + f\left(\frac{x}{k}\right) = f(x) \quad \text{for all } x \in \mathbb{R}_+, \tag{A.5}$$

which implies that

$$f\left(\frac{x}{k}\right) = \frac{1}{k}f(x) \quad \text{for all } x \in \mathbb{R}_+. \tag{A.6}$$

Combining (A.6) with (A.4), we have

$$f\left(\frac{j}{k}x\right) = \frac{j}{k}f(x) \quad \text{for all } x \in \mathbb{R}_+. \tag{A.7}$$

Notice further that

$$f\left(-\frac{j}{k}x\right) = -f\left(\frac{j}{k}x\right) = -\frac{j}{k}f(x) \quad \text{for all } x \in \mathbb{R}_+. \tag{A.8}$$

Since (A.8) holds for all positive integers $j$ and $k$, we have

$$f(yx) = yf(x) \quad \text{for all } x \in \mathbb{R}_+ \text{ and } y \in \mathbb{Q}. \tag{A.9}$$

Setting $x = 1$, we have

$$f(y) = yf(1) = \kappa y \quad \text{for all } y \in \mathbb{Q}. \tag{A.10}$$

Since the rational numbers are countable and dense in the real line and $f$ is continuous, Lemma 2 implies that

$$f(y) = \kappa y \quad \text{for all } y \in \mathbb{R}, \tag{A.11}$$

proving the result. □

**Lemma 4.** *Suppose that* $f : \mathbb{R} \to \mathbb{R}$ *is a continuous function satisfying* $f(0) = 0$ *and* $f(1) = 1$ *and suppose that*

$$f(x + yz) = f(x) + \big[f(x + y) - f(x)\big]f(z) \quad \textit{for all } (x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \tag{A.12}$$

*holds. Then* $f(x) = x$ *for all* $x \in \mathbb{R}$.

*Proof:* Plugging $x = 0$ into Equation (A.12) gives

$$f(yz) = f(y)f(z) \quad \text{for all } (y, z) \in \mathbb{R}_+ \times \mathbb{R}. \tag{A.13}$$

Plugging $z = x$ into (A.12) gives

$$f\big(x(y + 1)\big) = f(x) + \big[f(x + y) - f(x)\big]f(x) \quad \text{for all } (x, y) \in \mathbb{R} \times \mathbb{R}_+. \tag{A.14}$$

Using (A.13), we have $f(x(y + 1)) = f(x)f(y + 1)$, so that (A.14) gives

$$f(x)f(y + 1) = f(x) + \big[f(x + y) - f(x)\big]f(x) \quad \text{for all } (x, y) \in \mathbb{R} \times \mathbb{R}_+. \tag{A.15}$$

When $x \neq 0$, we can divide (A.15) by $f(x)$ to obtain

$$f(y + 1) = 1 + f(x + y) - f(x) \quad \text{for all } (x, y) \in \mathbb{R} - \{0\} \times \mathbb{R}_+. \tag{A.16}$$

Since $f$ is continuous, (A.16) must hold at $x = 0$ as well. We can therefore rearrange (A.16) to obtain

$$f(x + y) = f(x) + f(y + 1) - 1 \quad \text{for all } (x, y) \in \mathbb{R} \times \mathbb{R}_+. \tag{A.17}$$

Plugging $y = x = 1$ into (A.12), we obtain $f(z + 1) = 1 + [f(2) - 1]f(z)$ for all $z \in \mathbb{R}$, or equivalently,

$$f(y + 1) = 1 + \big[f(2) - 1\big]f(y) \quad \text{for all } y \in \mathbb{R}. \tag{A.18}$$

Plugging $x = 0$ and $y = 1$ into (A.17), we obtain

$$f(2) = 2. \tag{A.19}$$

Plugging (A.18) and (A.19) into (A.17), we obtain

$$f(x + y) = f(x) + f(y) \quad \text{for all } (x, y) \in \mathbb{R} \times \mathbb{R}_+. \tag{A.20}$$

By (A.20) and Lemma 3, we have $f(x) = x$ for all $x \in \mathbb{R}$, proving the result. □

**Lemma 5.** *Suppose that $f : A \cup B \to \mathbb{R}^K$ is continuous, where $A \subset \mathbb{R}^J$. Suppose that for all $\boldsymbol{x} \in B$ and $\varepsilon > 0$, there exists a $\boldsymbol{y} \in A$ such that $\|\boldsymbol{x} - \boldsymbol{y}\| < \varepsilon$. Suppose that $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in A$. Then $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in B$.*

*Proof:* Consider any $\boldsymbol{x} \in B$. Since $f$ is continuous at $\boldsymbol{x} \in B$, for each $\lambda > 0$, there exists an $\varepsilon > 0$ such that $\boldsymbol{y} \in A$ and $\|\boldsymbol{x} - \boldsymbol{y}\| < \varepsilon$ imply that $\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| < \lambda$. For each $\lambda > 0$, we can select $\boldsymbol{y} \in A$ with $\|\boldsymbol{x} - \boldsymbol{y}\| < \varepsilon$, so that $\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| < \lambda$. Since $f(\boldsymbol{y}) = 0$, we have $\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| = \|f(\boldsymbol{x})\| < \lambda$. Since this holds for all $\lambda > 0$, it follows that $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in B$, proving the result. □

## References


Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 392–479). Reading: Addison-Wesley.

Bock, R.D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, *6*, 443–459.

Bock, R.D., & Lieberman, M. (1970). Fitting a response curve model for dichotomously scored items. *Psychometrika*, *35*, 179–198.

Douglas, J.A. (1997). Joint consistency of nonparametric item characteristic curve and ability estimation. *Psychometrika*, *62*, 7–28.

Douglas, J.A. (2001). Asymptotic identifiability of nonparametric item response models. *Psychometrika*, *66*, 531–540.

Heckman, J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. In C. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 179–195). Cambridge: MIT Press.

Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, *27*, 887–906.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Mahwah: Lawrence Erlbaum Associates.

Peress, M., & Spirling, A. (2010). Scaling the critics: Uncovering the latent dimensions of movie criticism. *Journal of the American Statistical Association*, *105*, 71–83.

Poole, K.T. (2000). Non-parametric unfolding of binary choice data. *Political Analysis*, *8*, 211–237.

Poole, K.T., & Rosenthal, H. (1991). Patterns of congressional voting. *American Journal of Political Science*, *35*, 228–278.

Poole, K.T., & Rosenthal, H. (1997). *Congress: A political economic history of roll call voting*. New York: Oxford University Press.

Ramsay, J.O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, *56*, 611–630.

Ramsay, J.O., & Abrahamowicz, M. (1989). Binomial regression with monotone splines: A psychometric application. *Journal of the American Statistical Association*, *84*, 906–915.

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement*, *52*, 79–97.



*Manuscript Received: 6 MAR 2010*
*Final Version Received: 3 MAR 2011*