MULTIVARIATE BEHAVIORAL RESEARCH, *41*(4), 445–472

A Multidimensional Ideal Point Item Response Theory Model for Binary Data

Albert Maydeu-Olivares
Faculty of Psychology
University of Barcelona, and
Marketing Department
Instituto de Empresa

Adolfo Hernández
Department of Mathematical Sciences
University of Exeter

Roderick P. McDonald
Department of Psychology
University of Illinois

We introduce a multidimensional item response theory (IRT) model for binary data based on a proximity response mechanism. Under the model, a respondent at the mode of the item response function (IRF) endorses the item with probability one. The mode of the IRF is the ideal point, or in the multidimensional case, an ideal hyperplane. The model yields closed form expressions for the cell probabilities. We estimate and test the goodness of fit of the model using only information contained in the univariate and bivariate moments of the data. Also, we pit the new model against the multidimensional normal ogive model estimated using NOHARM in four applications involving (a) attitudes toward censorship, (b) satisfaction with life, (c) attitudes of morality and equality, and (d) political efficacy.

The normal PDF model is not invariant to simple operations such as reverse scoring. Thus, when there is no natural category to be modeled, as in many personality applications, it should be fit separately with and without reverse scoring for comparisons.

Correspondence concerning this article should be addressed to Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, P. Valle de Hebrón, 171. 08035 Barcelona, Spain. E-mail: amaydeu@ub.edu

INTRODUCTION

Over the past 30 years, item response theory (IRT) methods have enjoyed a growing popularity for modeling educational, personality, and attitudinal data. However, Chernyshenko, Stark, Chan, Drasgow, and Williams (2001, p. 524) argued that researchers and practitioners need to pay more attention to model-data fit when using IRT models. As they pointed out, without evidence of model fit, IRT results may be suspect. These authors compared the fit of Samejima's (1969) graded model and Levine's (1984) nonparametric Multilinear Formula Score (MFS) model to a number of personality scales and concluded that "traditional" parametric IRT models do not fit personality data well. They argued that, for personality data, ideal point models may be better suited than models based on a cumulative response process (such as Samejima's). As Chernyshenko et al. (p. 557) pointed out, IRT models based on a cumulative response process assume that the individual has a high probability of endorsing an item if the individual is located above the item on a joint scale as defined by Coombs (1964). In ideal point models the probability of an individual endorsing an item also depends on both the location of the person and the position of the statement along the latent trait continuum on such a joint scale. However, in these models, individuals will be more likely to agree with statements having scale values similar to their own, whereas they will be more likely to disagree with statements having scale values that are either more or less extreme. Thus, ideal point models have single-peaked item response functions. In the IRT literature ideal point models are commonly referred to as unfolding models. In this article, we use both terms interchangeably.

In a follow-up study to Chernyshenko et al. (2001), Maydeu-Olivares (2005) also compared the fit of a variety of parametric unidimensional models as well as Levine's model to the scales of another personality questionnaire. The methodology was similar to that of Chernyshenko et al. He concluded that among the parametric models considered, Samejima's was the best fitting model. However, in contrast to Chernyshenko et al.'s study, he found that in some situations Levine's model did not fit better than Samejima's model. Maydeu-Olivares attributed the discrepant results from the two studies to the amount of multidimensionality present in the scales of each study. He suggested that for scales that are substantially unidimensional, unidimensional IRT models with mononotic item response functions may be appropriate. On the other hand, for scales with moderate amounts of multidimensionality, multidimensional IRT models may be needed to yield an adequate fit.

The conclusions of Chernyshenko et al. (2001) and Maydeu-Olivares (2005) studies do not necessarily conflict. If an unfolding model is needed to provide an adequate fit to personality and attitudinal data as Chernyshenko et al. suggested, but there are non-negligible amounts of multidimensionality in the data to be modeled, then a multidimensional unfolding IRT model should be employed.

Ideal point models go at least as far back as Chave and Thurstone (1931), with fundamental contributions by Coombs (1964). By now there is a vast amount of literature on unfolding IRT models. To the interested reader van der Linden and Hambleton (1997, part 5) may be a useful starting point. At Professor James S. Roberts's Web page http://www.psychology.gatech.edu/unfolding/, readers also may find a brief introduction to these models, an exhaustive list of publications describing the numerous unidimensional unfolding IRT models, illustrative data sets, and information on programs for estimating some of these models. Yet to date there has been little research on multidimensional unfolding IRT models (but see Bradlow & Schmittlein, 2000; DeSarbo & Hoffman, 1986, 1987; Takane 1996, 1998). In fact, most research on multidimensional parametric IRT models has focused on a single model, the multidimensional extension of Samejima's model, using either a logistic function or a normal ogive function. In the multidimensional case, the normal ogive version of the model is most frequently used, and the model is simply referred to as the multidimensional normal ogive model (e.g., McDonald, 1997).

In this article we introduce a new ideal point IRT model for binary data. When a single latent trait is assumed to underlie the responses, the model assumes that there is an ideal point for each item, the maximum of the item response function, and that when a respondent's position on the latent trait continuum coincides with the ideal point, the respondent will endorse the item with probability one. Thus, the model assumes that there is truly an "ideal point" for each item. The greater the distance between the respondent's position and the ideal point, the smaller the probability of endorsing the item. Our model is multidimensional in that the response to an item may depend on more than one latent trait. In that case, our model leads to an ideal hyperplane rather than to an ideal point. Thus for instance, in two-dimensional models our model invokes the existence of an ideal line in the two-dimensional space.

With the aim to explore model-data fit in personality and attitudinal data, and in particular in relation to the choice of cumulative versus noncumulative item response functions, we pit our model against the multidimensional normal ogive model in a series of applications. The normal ogive model is obtained by using a normal distribution function to link the conditional probability of endorsing an item to a linear function of the latent traits. In the model proposed here we simply use a normal probability density function (PDF) as link function instead of the cumulative distribution function. Accordingly, we use the term *normal PDF model* to denote the model introduced here. Because in addition the latent traits are assumed to be normally distributed, our model captures the notion of a proximity response mechanism through the use of the normal density function twice. First, the normal density function is used to model the density of the respondents' latent traits. Second, this function is used to model the conditional probability of endorsing an item given the latent traits (i.e., the item response function).

The most frequently used method for estimating the multidimensional normal ogive model is probably the limited information procedure implemented in NOHARM (Fraser & McDonald, 1988). In NOHARM consistent and asymptotically normal estimates are obtained by simply minimizing an unweighted least squares discrepancy between sample and model-implied univariate and bivariate moments of the data. A similar procedure is used to estimate the model introduced here.

THE NORMAL PDF MODEL

Consider n items Y = $(Y_1, \ldots, Y_n)'$ each with two possible outcomes. Without loss of generality we may assign the values {0, 1} to their outcomes. Thus, the distribution of each item Y_i is Bernoulli, and the joint distribution of the set of items **Y** is multivariate Bernoulli (MVB: Teugels, 1990; Maydeu-Olivares & Joe, 2005).

Any item response model for this set of items can be written as (Bartholomew & Knott, 1999)

$$\Pr\left(\bigcap_{i=1}^{n}(Y_i = y_i)\right) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\prod_{i=1}^{n}\left[\Pr\left(Y_i = 1|\boldsymbol{\eta}\right)\right]^{y_i}\left[1-\Pr\left(Y_i = 1|\boldsymbol{\eta}\right)\right]^{1-y_i}\right\}\gamma_p\left(\boldsymbol{\eta}\right)d\boldsymbol{\eta} \quad (1)$$

Here, we use $\Pr\left(\bigcap_{i=1}^{n}(Y_i = y_i)\right)$ to denote the probability of observing each of the possible 2^n binary patterns, and $\gamma_p(\boldsymbol{\eta})$ to denote the probability density function of a p-dimensional vector of latent traits $\boldsymbol{\eta}$. Finally, $\Pr(Y_i = 1|\boldsymbol{\eta})$ is denoted in the IRT literature the item response function. Different IRT models can be obtained by selecting suitable models for $\Pr(Y_i = 1|\boldsymbol{\eta})$ and $\gamma_p(\boldsymbol{\eta})$ in (1). For a good overview of IRT models, see van der Linden and Hambleton (1997). However, due to the difficulty in evaluating the multidimensional integral in (1) most latent trait models proposed to date assume a single latent trait. Obviously, unidimensional latent trait models (i.e., $p = 1$) are less likely to be able to yield a good fit in applications than multidimensional models unless applied to a well-designed homogeneous item set.

The most widely used multidimensional latent trait model is the normal ogive model (e.g., McDonald, 1997). This model assumes that the item response function is a standard normal distribution function evaluated at $\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta}$, that is

$$\Pr\left(Y_i = 1|\boldsymbol{\eta}\right) = \Phi_1\left(\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta}\right), \quad (2)$$

where $\boldsymbol{\beta}_i' = \left(\beta_{i1}, \cdots, \beta_{ip}\right)$. It also assumes that the density of the latent traits is multivariate standard normal with correlation matrix $\boldsymbol{\Psi}$, that is,

$$\gamma_p(\boldsymbol{\eta}) = \phi_p(\boldsymbol{\eta} : \mathbf{0}, \boldsymbol{\Psi}). \tag{3}$$

Like the normal ogive model, most latent trait models for binary data use monotonically increasing item response functions. However, some authors (e.g., Andrich, 1996; Roberts, 1995; van Schuur & Kiers, 1994) have argued, following Coombs (1964), that if the psychological mechanism by which individuals respond to items is a proximity mechanism, nonmonotonic item response functions should be used instead. Models based on this assumption of a proximity psychological response mechanism are generally referred to as ideal point models and as unfolding models.

To obtain a multidimensional unfolding model we assume as in the normal ogive model that Equation (3) holds. That is, we assume that the latent traits are normally distributed, and possibly correlated. However, the item response function is specified as

$$\Pr(Y_i = 1|\boldsymbol{\eta}) = \sqrt{2\pi}\phi_1(\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta} : 0,1) = \exp\left[-\frac{1}{2}(\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta})^2\right]. \tag{4}$$

Namely, we use a normal density function (scaled by $\sqrt{2\pi}$) to link the item response function to the linear function of the latent traits $\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta}$. We refer to this model as the normal PDF model. The constant $\sqrt{2\pi}$ is used to ensure that the item response function takes the value of one for some value of the latent traits $\boldsymbol{\eta}$. Thus, according to the model if the respondent and item positions coincide in the space of the latent traits, the respondent will endorse the item with probability one.

In the case of a single latent trait (i.e., $p = 1$), the nonmonotonic and symmetric item response function of this model reaches its maximum at $\eta = -\frac{\alpha_i}{\beta_i}$ and has two inflexion points $\eta = \frac{\pm 1 - \alpha_i}{\beta_i}$. That is, the ideal point is $\mathrm{H}_i = -\alpha_i/\beta_i$. In multidimensional models, we have an ideal hyperplane rather than an ideal point. This is because the maximum of (4) is reached whenever $H_i : \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta} = 0$. Using the general expression of the squared Euclidean distance between a point and a hyperplane,

$$d_E^2(\boldsymbol{\eta}, H_i) = \frac{1}{\boldsymbol{\beta}_i'\boldsymbol{\beta}_i}(\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\eta})^2, \tag{5}$$

we can write the item response function of the normal PDF model as a function of this distance as

$$\Pr(Y_i = 1|\boldsymbol{\eta}) = \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_i'\boldsymbol{\beta}_i) \times d_E^2(\boldsymbol{\eta}, H_i)\right]. \tag{6}$$

Note that in the unidimensional case, $d_E^2\left(\eta,-\frac{\alpha_i}{\beta_i}\right)=\left(\eta+\frac{\alpha_i}{\beta_i}\right)^2=\frac{1}{\beta_i^2}(\alpha_i+\beta_i\eta)^2$, and we obtain the special case $\Pr\left(Y_i=1|\eta\right)=\exp\left[-\frac{1}{2}\beta_i^2\times d_E^2\left(\eta,-\frac{\alpha_i}{\beta_i}\right)\right]$.

We also note that in general, for $p \geq 1$, (4) satisfies

$$\sqrt{2\pi}\phi_1\left(\alpha_i+\boldsymbol{\beta}_i'\boldsymbol{\eta}\right)=\sqrt{2\pi}\phi_1\left(-\alpha_i-\boldsymbol{\beta}_i'\boldsymbol{\eta}\right)$$

and

$$\sqrt{2\pi}\phi_1\left(-\alpha_i+\boldsymbol{\beta}_i'\boldsymbol{\eta}\right)=\sqrt{2\pi}\phi_1\left(\alpha_i-\boldsymbol{\beta}_i'\boldsymbol{\eta}\right). \tag{7}$$

To resolve this indeterminacy we estimate the model parameters with the restriction that the intercepts $\boldsymbol{\alpha}=\left(\alpha_1,\cdots,\alpha_n\right)'$ be negative.

The normal PDF model satisfies another very interesting property, namely, that the probability of the full vector of responses of an individual can be all computed in closed form, without requiring integration, regardless of the dimensionality of the latent traits. In other words, for this model, given a set of item parameter estimates, Equation (1) has a closed form solution. To see this, let $\mathbf{B}=(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)'$ be a $n \times p$ matrix of slope parameters, and let $\boldsymbol{\Sigma}=\mathbf{I}_n+\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$. Furthermore, let $\mathbf{s}$ be any subset consisting of k items. Because the distribution of the items is Bernoulli, we have $\mu_s=E\left[\prod_{i\in\mathbf{s}}Y_i\right]=\Pr\left[\bigcap_{i\in\mathbf{s}}(Y_i=1)\right]$ (Teugels, 1990). That is, μ_s is the kth joint moment involving the variables in $\mathbf{s}$. It is also the probability that all the variables in $\mathbf{s}$ take the value of 1. We show in the Appendix that μ_s has the following closed form solution under the normal PDF model,

$$\begin{aligned}\mu_\mathbf{s}&=\Pr\left[\bigcap_{i\in\mathbf{s}}(Y_i=1)\right]=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\prod_{i\in\mathbf{s}}\varphi\left(\alpha_i+\boldsymbol{\beta}_i'\boldsymbol{\eta}\right)\right]\phi_p\left(\boldsymbol{\eta}:\mathbf{0},\boldsymbol{\Psi}\right)d\boldsymbol{\eta}=\\&=\left(2\pi\right)^{\frac{k}{2}}\phi_1\left(\alpha_\mathbf{s}:\mathbf{0},\boldsymbol{\Sigma}_\mathrm{s}\right),\end{aligned} \tag{8}$$

where $\boldsymbol{\alpha}_\mathrm{s}$ and $\boldsymbol{\Sigma}_\mathrm{s}$ denote a $k \times 1$ vector and a $k \times k$ matrix, respectively, obtained by taking the appropriate rows and columns of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$.

For instance, applying (8) we find that the univariate and bivariate moments of the MVB distribution under the normal PDF model are

$$\pi_i=E[Y_i]=\Pr(Y_i=1)=\sqrt{2\pi}\phi_1\left(\alpha_i:0,1+\boldsymbol{\beta}_i'\boldsymbol{\beta}_i\right) \tag{9}$$

$$\pi_{ij} = E[Y_i Y_j] = \Pr[(Y_i = 1) \cap (Y_j = 1)] = 2\pi\phi_2(\boldsymbol{\alpha}_{ij} : \mathbf{0}, \mathbf{I} + \mathbf{B}_{ij}{}' \boldsymbol{\Psi} \mathbf{B}_{ij}) \tag{10}$$

where $\boldsymbol{\alpha}_{ij} = \left(\alpha_i, \alpha_j\right)'$ and $\mathbf{B}_{ij} = (\mathbf{B}_i \mid \mathbf{B}_j)$ is a $p \times 2$ matrix. We used $\pi_i = \mu_i$ and $\pi_{ij} = \mu_{ij}$ in (9) and (10) to indicate that, because the items are Bernoulli random variables, these moments are also marginal probabilities. The full set of $(2^n - 1)$ moments of the MVB distribution can be computed in closed form using (8). Because there is a one-to-one linear correspondence between the set of these moments and the set of 2^n binary cell probabilities (Maydeu-Olivares & Joe, 2005; Teugels, 1990), the cell probabilities (1) can be computed also in closed form.

To identify the model it suffices to consider univariate and bivariate MVB moments. Identification restrictions remain unchanged when higher order moments are considered. This implies that the model can be estimated using only the univariate and bivariate margins of the contingency table. It also implies that the identification conditions for the normal PDF model are identical to those in the normal ogive model. Identification conditions for the normal ogive model are given for instance in McDonald (1999). Thus, when $p = 1$, all parameters of the normal PDF model are identified. When $p > 1$ a model with minimal identification restrictions (i.e., an unrestricted or exploratory model) is obtained by setting $\boldsymbol{\Psi} = \mathbf{I}$ and, to solve the rotational indeterminacy of $\mathbf{B}$, by letting $\mathbf{B}$ be a low echelon matrix (i.e., $\beta_{hl} = 0$, $l = 1,\ldots, p$; $h = 1,\ldots, l - 1$). In the multidimensional case, after the item parameters have been estimated, $\hat{\mathbf{B}}$ may be rotated orthogonally or obliquely to help in interpreting the model, just as in the normal ogive model. Alternatively, based on some a priori information about the data, researchers may wish to fit a restricted model in which some elements of $\boldsymbol{\alpha}$, $\mathbf{B}$, and $\boldsymbol{\Psi}$ are subject to normalization, exclusion or equality constraints.

In closing our treatment of the normal PDF IRT model we consider making statements about an individual's location on the latent traits given his or her binary responses. All the relevant information needed for this is contained in the posterior distribution of the latent traits given the observed binary responses (Bartholomew & Knott, 1999),

$$\varphi_p\left(\boldsymbol{\eta} \middle| \mathbf{y}\right) = \frac{\gamma_p\left(\boldsymbol{\eta}\right)\left\{\prod_{i=1}^{n}\left[\Pr\left(Y_i = 1\middle|\boldsymbol{\eta}\right)\right]^{y_i}\left[1 - \Pr\left(Y_i = 1\middle|\boldsymbol{\eta}\right)\right]^{1-y_i}\right\}}{\Pr\left(\bigcap_{i=1}^{n} Y_i = y_i\right)}. \tag{11}$$

Thus, after the item parameters have been estimated, an individual's location can be obtained for instance by computing the mean or the mode of this posterior distribution. The former are known as expected a posteriori (EAP) scores, and the latter maximum a posteriori (MAP) scores. Obtaining MAP scores in general requires an iterative procedure, whereas obtaining EAP scores involves computing

$$EAP(\mathbf{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{\eta}\varphi_p(\boldsymbol{\eta}|\mathbf{y})\,d\boldsymbol{\eta}. \quad (12)$$

We illustrate the use of EAP scores in one of the applications.

LIMITED INFORMATION ESTIMATION AND TESTING

Because MVB moments have a closed form solution under this model, estimation methods that minimize a discrepancy function between sample and expected moments are a natural choice.[1] However, in most latent trait applications the number of binary variables is large and the observed contingency tables are very sparse. As a result, high order sample moments may be very poorly estimated. In contrast, univariate and bivariate sample moments can be reasonably estimated in very small samples regardless of n. Limited information procedures based on univariate and bivariate information are the most widely used approaches to estimate the multidimensional normal ogive model (see Christoffersson, 1975; Maydeu-Olivares, 2001b; Muthén, 1978, 1993). These methods also yield as a side product limited information goodness-of-fit tests. Maydeu-Olivares and Joe (2005) recently provided a unified treatment of limited and full information estimation and goodness-of-fit testing methods for MVB models. They show that bivariate information methods have high efficiency relative to asymptotically optimal procedures such as maximum likelihood. Furthermore, they are considerably faster to execute, particularly for multidimensional models. They also show that bivariate information tests have more precise Type I errors and are asymptotically more powerful in large and sparse binary tables than full information goodness-of-fit tests such as Pearson's χ^2.

[1]The model can be alternatively estimated using full information maximum likelihood (ML). However, when the normal PDF model is estimated using ML the computational advantage of this model (closed form expressions for the MVB moments) is lost for large models. This is because, in order to compute the binary pattern probabilities using the closed form expressions for the moments, all $2^n - 1$ moments need to be computed. As the number of variables increases, $2^n - 1$ becomes a very large number. For ML estimation of large models it is computationally more efficient to compute the probabilities of the *observed* patterns by numerical integration than computing the probabilities of *all* patterns using the closed form expressions for the MVB moments.

Limited information methods using only univariate and bivariate moments are employed here to estimate and evaluate the goodness of fit of the normal PDF model. To estimate the model, we collect all n univariate moments given by Equation (9) in $\dot{\boldsymbol{\pi}}_1$ and all $\frac{n(n-1)}{2}$ bivariate moments given by Equation (10) in $\dot{\boldsymbol{\pi}}_2$. Thus, we let $\dot{\boldsymbol{\pi}}' = (\dot{\boldsymbol{\pi}}_1{}', \dot{\boldsymbol{\pi}}_2{}')$, where $\dot{\boldsymbol{\pi}}_1{}' = \left(\pi_1, \cdots, \pi_n\right)$ and $\dot{\boldsymbol{\pi}}_2{}' = (\pi_{12}, \ldots, \pi_{1n}, \pi_{23}, \cdots, \pi_{2n}, \cdots \pi_{n-1,n})$. We use $\dot{\boldsymbol{\pi}}(\boldsymbol{\theta})$ to denote the restrictions imposed by the model on the vector of univariate and bivariate moments $\dot{\boldsymbol{\pi}}$, where $\boldsymbol{\theta}$ denotes a q-dimensional vector containing all mathematically independent elements in $\boldsymbol{\alpha}$, $\mathbf{B}$, and $\boldsymbol{\Psi}$. Thus, the degrees of freedom available for testing are $r = \frac{n(n+1)}{2} - q$. We assume that $\boldsymbol{\Delta} = \frac{\partial \dot{\boldsymbol{\pi}}}{\partial \boldsymbol{\theta}'}$ is of full rank so that the model is locally identified. Furthermore, let N denote sample size and let $\dot{\mathbf{p}}$ be the sample counterpart of $\dot{\boldsymbol{\pi}}$ (i.e., univariate and bivariate sample proportions). Then, the model parameters can be estimated by minimizing

$$F = \mathbf{e}'\hat{\mathbf{W}}\mathbf{e}. \tag{13}$$

where $\mathbf{e} = \left(\dot{\mathbf{p}} - \dot{\boldsymbol{\pi}}(\boldsymbol{\theta})\right)$, and $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a positive definite matrix. Letting $\boldsymbol{\Gamma}$ be the asymptotic covariance matrix of $\sqrt{N}\dot{\mathbf{p}}$, some common choices of $\hat{\mathbf{W}}$ in (13) are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Gamma}}^{-1}$ (weighted least squares, or WLS), $\hat{\mathbf{W}} = \left(\mathrm{Diag}\left(\hat{\boldsymbol{\Gamma}}\right)\right)^{-1}$ (diagonally weighted least squares, or DWLS), and $\hat{\mathbf{W}} = \mathbf{I}$ (unweighted least squares, or ULS).

This general estimation framework is denoted as weighted least squares for moment structures (see Browne, 1984; Browne & Arminger, 1995; Satorra & Bentler, 1994). Maydeu-Olivares and Joe (2005; see also Maydeu-Olivares, 2001b) provided a unified framework of full and limited information weighted least squares estimation methods for MVB models. Large sample properties for the parameter estimates, standard errors and goodness-of-fit tests of the model can be readily obtained using standard theory for the estimation of moment structures. Letting $\mathbf{H} = \left(\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}'\mathbf{W}$, the estimator $\hat{\boldsymbol{\theta}}$ obtained by minimizing (13) is consistent and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}\boldsymbol{\Gamma}\mathbf{H}'\right) \tag{14}$$

$$\sqrt{N}\hat{\mathbf{e}} \xrightarrow{d} N\left(\mathbf{0}, \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\Gamma}\left(\mathbf{I} - \mathbf{H}\right)\right). \tag{15}$$

where $\hat{\mathbf{e}} = (\dot{\mathbf{p}} - \dot{\boldsymbol{\pi}}(\hat{\boldsymbol{\theta}}))$ denotes the univariate and bivariate residuals. These residuals can be divided by their standard error to obtain standardized residuals that are asymptotically standard normal. Also, in unrestricted multidimensional solutions where the columns of $\hat{\mathbf{B}}$ have been rotated, standard errors for the rotated loadings can be obtained using formulae given by Maydeu-Olivares (2001b); see also Browne and du Toit (1992).

Now, from standard theory, $T := N\hat{F} \stackrel{a}{=} \mathbf{e}'\mathbf{W}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{e}$, where $\stackrel{a}{=}$ denotes asymptotic equality. Thus, in general,

$$T \xrightarrow{d} \sum_{i=1}^{r} \lambda_i \chi^2, \tag{16}$$

where the χ^2's are independent chi-square variables with one degree of freedom and the λ_i's are the non-null eigenvalues of $\mathbf{M} = \mathbf{W}\left(\mathbf{I}-\mathbf{\Delta H}\right)\mathbf{\Gamma}$. In particular, when $\hat{\mathbf{W}} = \hat{\mathbf{\Gamma}}^{-1}$, we obtain an estimator with minimum asymptotic variance among the class of estimators (13). In this special case (14) and (16) simplify to $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0},(\mathbf{\Delta\Gamma}^{-1}\mathbf{\Delta})^{-1})$ and $T \xrightarrow{d} \chi^2_r$. respectively. However, the use of $\hat{\mathbf{W}} = \hat{\mathbf{\Gamma}}^{-1}$ requires inverting a very large matrix. Thus, WLS estimation is not suitable for large applications.

When $\hat{\mathbf{W}} \neq \hat{\mathbf{\Gamma}}^{-1}$, following Satorra and Bentler (1994; see also Maydeu-Olivares, 2001a, 2001b; Rao & Scott, 1987) to assess the goodness of fit of the model we may scale T by its asymptotic mean using

$$\bar{T} = \frac{T}{c}, \qquad c = \frac{\mathrm{Tr}\left(\mathbf{M}\right)}{r}. \tag{17}$$

Alternatively, we may adjust T by its asymptotic mean and variance using

$$\bar{\bar{T}} = \frac{T}{\mathrm{Tr}\left(\mathbf{M}^2\right)/r}. \tag{18}$$

$\bar{T}$ and $\bar{\bar{T}}$ denote the scaled (for mean) and adjusted (for mean and variance) test statistics. The former is referred to a chi-square distribution with r degrees of freedom, whereas the latter is referred to a chi-square distribution with $d = \frac{\mathrm{Tr}\left(\mathbf{M}\right)^2}{\mathrm{Tr}\left(\mathbf{M}^2\right)/r}$ degrees of freedom.

Here we estimate the normal PDF model by simply using $\hat{\mathbf{W}} = \mathbf{I}$ (i.e., ULS) where standard errors, standardized residuals, and goodness-of-fit tests will be computed via (14), (15), and (17) by evaluating $\mathbf{\Delta}$ and $\mathbf{\Gamma}$ at the estimated parameter values. This approach is very similar to the one implemented in the computer program NOHARM (Fraser & McDonald, 1988) which estimates the multidimensional normal ogive model also using ULS from univariate and bivariate moments.

However, there are two differences between the present approach and the approach used in NOHARM (see Maydeu-Olivares, 2001b). The first difference is that in NOHARM estimation is performed in two stages to improve computational efficiency by exploiting a separability of parameters that exists in the normal ogive model but not in the normal PDF model. The second difference is that to obtain standard errors and goodness-of-fit tests for NOHARM, $\mathbf{\Gamma}$ is consistently estimated using sample proportions, whereas in our estimation of the normal PDF model $\mathbf{\Gamma}$ is consistently estimated by evaluating it at the estimated parameter values. This is because $\mathbf{\Gamma}$ depends on fourth-order joint moments. These can be computed in closed form under the normal PDF model, but they require multivariate integration in the normal ogive model.

APPLICATIONS

We present four applications where we compare the fit of the normal PDF model estimated using the ULS estimator described above against the fit of the normal ogive model estimated using also ULS as implemented in NOHARM. In the first application, we model attitudes toward censorship. This is a typical application where the item stems suggest that it is plausible to assume that individuals use a proximity mechanism in responding to the items. In applications like this one, the normal PDF model should provide a better fit than the normal ogive model. In the second application we model satisfaction with life. In this application, given the item stems, a proximity mechanism in responding to the items does not seem plausible. Rather, a priori, a model with monotonically increasing item response functions seems more reasonable for these items. As a result, in this application we expect the normal ogive model to yield a better fit. The third application involves modeling attitudes of morality and equality. This application is used to illustrate that if the sample size is small, the normal PDF and normal ogive model may yield a similar fit although the estimated item response functions appear quite distinct. Finally, the fourth application involves modeling political efficacy. This application is used to illustrate that in some situations we may not empirically distinguish between these two models, even with large sample sizes, because their item response functions are very similar in the region of high density of respondents.

Attitudes Toward Censorship

In this example, we model a set of 223 observations collected by Roberts (1995) on 20 statements reflecting attitudes toward censorship.[2] The statements were originally published in Rosander and Thurstone (1931). Roberts (1995) asked the re-

[2]The data are available at http://www.psychology.gatech.edu/unfolding/data.html.

TABLE 1
Goodness-of-Fit Tests for the Censorship Data

Items	*Traits*	T	$\bar{T}$	*df*	p	$\bar{\bar{T}}$	*df*	p
Normal PDF								
20	1	5.40	244.38	170	< .01	183.40	127.58	< .01
20	2	3.58	180.15	151	.05	126.86	106.33	.09
20	3	2.66	145.42	133	.22	104.00	95.12	.25
16	1	2.59	133.64	104	.03	95.92	74.65	.05
Normal ogive								
20	1	9.73	408.50	170	< .01	165.05	68.67	< .01
20	2	4.27	214.21	151	< .01	99.30	70.00	.01
20	3	3.07	169.77	133	.02	83.83	65.67	.07
16	1	3.63	171.01	104	< .01	85.98	52.29	< .01

Note. $N = 223$; $T = N\hat{F}$; $\bar{T}$ denotes T adjusted by its asymptotic mean; $\bar{\bar{T}}$ denotes T adjusted by its asymptotic mean and variance.

spondents to rate each statement using a 6-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Their responses were dichotomized (0 = disagree, 1 = agree) for this analysis.

In Table 1 we provide goodness-of-fit results for the normal PDF model applied to these data with one, two, and three latent traits.[3] In this table, we also provide the goodness of fit results obtained for the normal ogive model using NOHARM. As can be seen in this table, the normal PDF model reproduces the data better than the normal ogive model. Two latent trait dimensions seem to be necessary to reproduce these data using the normal PDF model, whereas three dimensions seem to be necessary for the normal ogive model.

In Figure 1 we provide a plot of the slope parameters estimated in the two-dimensional solution, that is, $\hat{\mathbf{B}}$. Also, in Figure 2 we provide the bidimensional item response function for one of the items, Item 5. We note that in this figure the ideal point (i.e., the point at which respondents endorse an item with probability one) is not a point but an hyperplane. In this case, because the model is bidimensional, the hyperplane is just a line.

Now, in Figure 1, most items fall roughly on a straight line close to the latent trait 1 axis. High scores on this axis indicate an anti-censorship attitude, and low scores on this axis a pro-censorship attitude. We cannot meaningfully interpret the second latent trait. In fact, a mild (20º) rotation suggests that the second dimension is mostly caused by Item 5. Thus, the second latent trait appears to be just "noise" induced by some items which are not appropriate indicators of the first latent trait.

[3]The null hypothesis being tested is that the model holds in the population. Larger values on the statistics indicate worse fit. As a result, the statistics could be referred to as "badness-of-fit" statistics. Here we use of the more common term *goodness-of-fit statistics.*

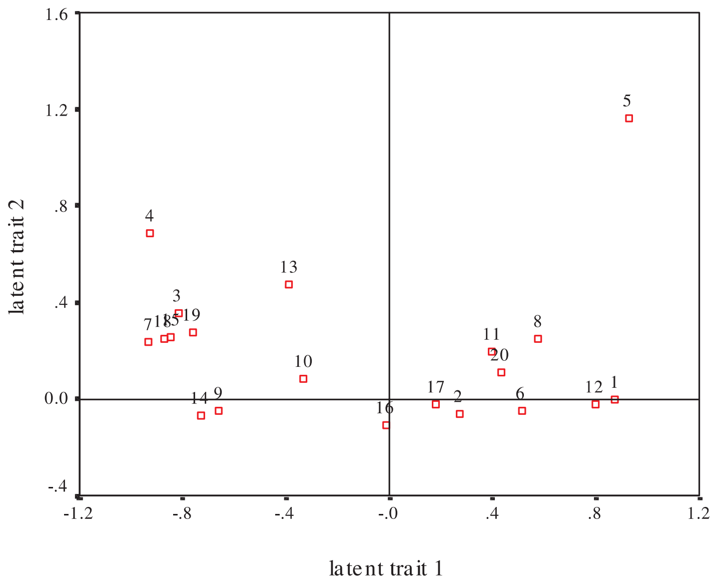


FIGURE 1 Censorship data. Plot of the regression slopes for the two-dimensional solution.

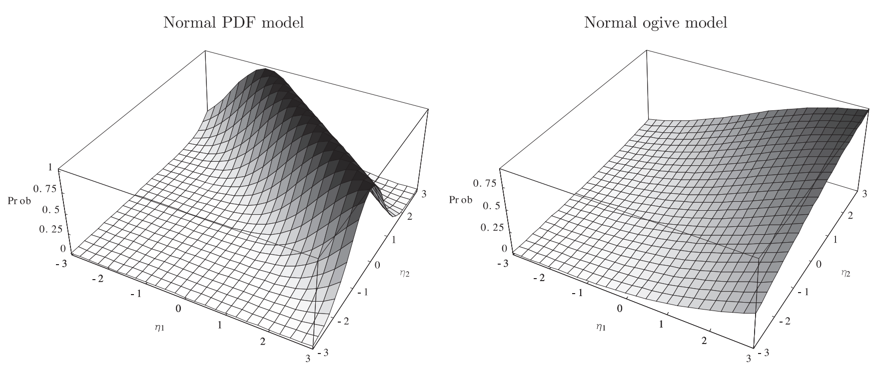


FIGURE 2 Plot of the item response function Pr $(Y = 1|\eta)$ for item 5 of the censorship data.

To identify which items are poor indicators of the latent trait *attitude toward censorship*, we inspect the standardized univariate and bivariate residuals obtained from fitting the one-dimensional model. The five largest standardized residuals for the one-dimensional solution are {–6.33, –4.84, 3.64, –3.14, –2.96} which correspond to the following univariate and bivariate residuals {(4), (12), (19,4), (5), (4,1)}. These residuals suggest that a one-dimensional solution may fit the 17 statements remaining after deleting statements {4, 5, 12}. The goodness-of-fit indices for this model are $\overline{T}$ = 145.01 on 119 *df*, *p* = .05, and $\overline{\overline{T}}$ = 105.30 on 86.41 *df*, *p* = .08. This model fits well. Interestingly, the β estimate for Item 16 is very low, 0.01, with a standard error of 0.06. Under a unidimensional normal PDF model this statement "Education of the public taste is preferable to censorship" provides very little information about respondents' attitudes toward censorship. After removing this item, we fitted a one-dimensional model to the remaining 16 items obtaining finally $\overline{T}$ = 133.64 on 104 *df*, *p* = .03, and $\overline{\overline{T}}$ = 95.92 on 74.65 *df*, *p* = .05. Thus, we have been able to identify a set of 16 items from the original set that can be used to measure attitude toward censorship. The parameter estimates and standard errors for this final model are given in Table 2.

TABLE 2
Parameter Estimates and Standard Errors for a One-Dimensional Model Applied to 16 Statements on Censorship

Item	*Stem*	α	β
1	I doubt if censorship is wise.	–1.20 (0.10)	–0.87 (0.24)
2	A truly free people must be allowed to choose their own reading and entertainment.	–0.28 (0.10)	–0.33 (0.10)
3	We must have censorship to protect the morals of young people.	–1.05 (0.09)	0.77 (0.16)
6	The whole theory of censorship is utterly unreasonable.	–1.54 (0.08)	–0.50 (0.10)
7	Until public taste has been educated, we must continue to have censorship.	–1.48 (0.11)	0.98 (0.22)
8	Many of our greatest literary classics would be suppressed if the censors thought they could get away with it.	–0.63 (0.06)	–0.27 (0.07)
9	Everything that is printed for publication should first be examined by government censors.	–1.91 (0.11)	0.63 (0.13)
10	Plays and movies should be censored but the press should be free.	–1.88 (0.09)	0.30 (0.10)
11	Censorship has practically no effect on people's morals.	–1.36 (0.07)	–0.33 (0.08)
13	Censorship protects those who lack judgment or experience to choose for themselves.	–1.02 (0.06)	0.38 (0.08)
14	Censorship is a very difficult problem and I am not sure how far I think it should go.	–0.16 (0.10)	0.76 (0.07)
15	Censorship is a good thing on the whole although it is often abused.	–0.75 (0.11)	0.88 (0.14)
17	Human progress demands free speech and a free press.	–0.25 (0.09)	–0.43 (0.08)
18	Censorship is effective in raising moral and aesthetic standards.	–1.13 (0.10)	0.89 (0.18)
19	Censorship might be warranted if we could get reasonable censors.	–0.47 (0.10)	0.80 (0.10)
20	Morality is produced by self-control, not by censorship.	–0.30 (0.10)	–0.42 (0.09)

Note. The statement numbering corresponds to the original 20-item set; standard errors are in parentheses.

We provide in Figure 3 plots of the item response function for selected items. The plots in Figure 3 illustrate the versatility of the model. The item response function for Item 9 corresponds to an item that is endorsed only by respondents with a pro-censorship view. The probability of endorsing this item is maximum for extreme pro-censorship respondents and the item response function in the region of high density of respondents is monotonically increasing. The plot for Item 3 ("We must have censorship to protect the morals of young people") on the other hand is non-monotonic. According to the model, the probability of endorsing the item is maximum for respondents with a moderately positive attitude toward censorship. The more anti-censorship the attitude, the less likely the item is endorsed. But respondents with an extreme pro-censorship view are also less likely to endorse the item than respondents with an moderate pro-censorship view. They may not endorse the item because they believe that the morals of all people should be protected, not only the morals of young people. Finally, the model can handle well "I do not have a clear opinion on the topic" items such as Item 14. The probability of endorsing this item is maximum for respondents that are neither pro- nor anti-censorship and minimum for respondents with extreme pro- or anti-censorship views.

Given this discussion, it is not surprising that, as shown in Table 1, a unidimensional model with monotonically increasing item response functions, such as the normal ogive model, fails to fit adequately these 16 items.

Satisfaction With Life

Professor Edward Diener kindly provided the responses of 7,167 individuals from 42 countries to the Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985). The questionnaire consists of these five items.

1. In most ways my life is close to my ideal.
2. The conditions of my life are excellent.
3. I am satisfied with my life.
4. So far I have gotten the important things I want in life.
5. If I could live my life over, I would change almost nothing.

which are to be rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). For this analysis we discarded those individuals who chose the middle category *neither agree nor disagree* for any of the items and dichotomized the responses of the remaining individuals (0 = *disagree,* 1 = *agree*). The resulting sample size was 4,073.

Of these item stems only the first and third may be consistent with the notion of a proximity response mechanism. Thus, a priori, we expect a model with monotonic curves to fit better these data than the normal PDF model. In Table 3 we provide goodness-of-fit results for one- and two-dimensional normal PDF and normal ogive models fitted to these data. We also include in this table the results of a re-

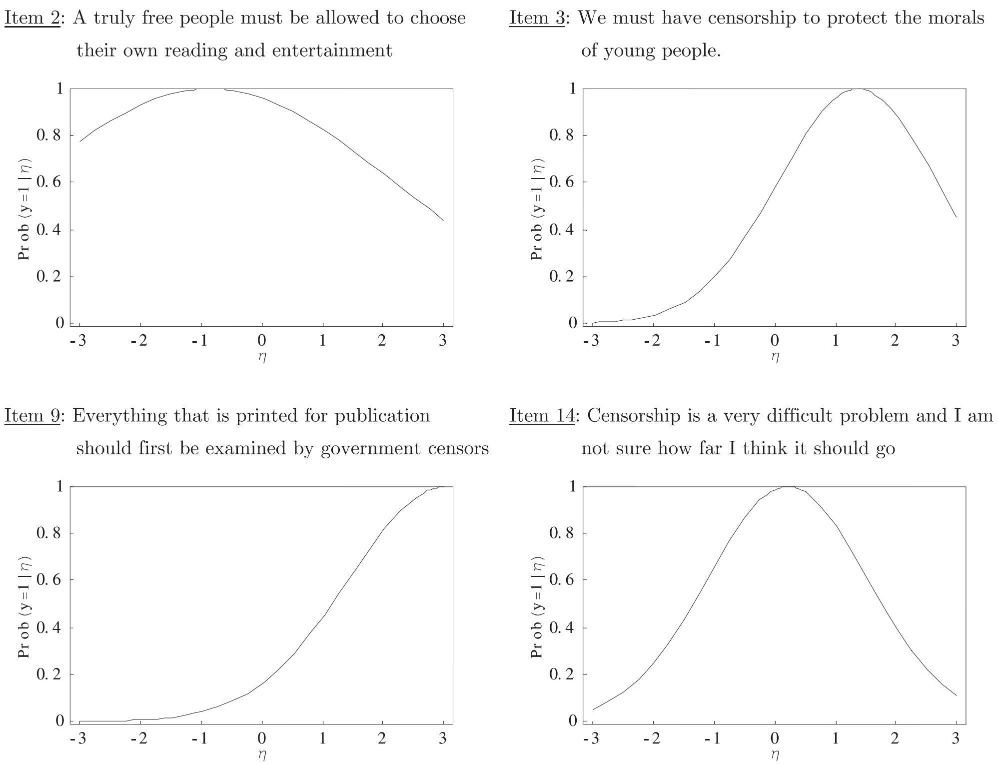


FIGURE 3 Plot of selected item response functions for a one dimensional model applied to 16 items of the censorship data.

TABLE 3
Goodness-of-Fit Tests for the Satisfaction With Life Scale

	T	$\bar{T}$	*df*	*p*	$\bar{\bar{T}}$	*df*	*p*
Normal PDF							
One trait	7.24	273.70	5	< .01	258.53	4.72	< .01
Two traits, unrestricted	7.18	198.79	1	< .01	198.79	1.00	< .01
Two traits, restricted	6.37	358.19	4	< .01	178.16	1.79	< .01
Normal ogive							
One trait	0.93	42.84	5	< .01	41.14	4.80	< .01
Two traits, unrestricted	0.01	0.75	1	.39	0.75	1.00	.39
Two traits, restricted	0.21	10.20	4	.04	10.04	3.94	.04

Note. N = 4,073; $\bar{T}$ denotes T adjusted by its asymptotic mean; $\bar{\bar{T}}$ denotes T adjusted by its asymptotic mean and variance.

stricted two-dimensional model suggested by McDonald (1999). In this model, the first three items are indicators of present satisfaction with life, the two last items measure past satisfaction with life, and the two latent dimensions are correlated. As can be seen in this table, the two-dimensional restricted normal ogive model provides a good fit to these data given the sample size. On the other hand, all the normal PDF models provide an extremely poor fit to these data. This was what we expected given the item stems.

Attitudes of Morality and Equality

Jöreskog and Sörbom (1996) provided data on 200 Swedish schoolchildren in Grade 9 who used a 4-point scale (*unimportant*, *not important*, *important*, and *very important*) to rate the importance of each of these items to them: (a) human rights, (b) equal conditions for all people, (c) racial problems, (d) equal value of all people, (e) euthanasia, (f) crime and punishment, (g) conscientious objectors, and (h) guilt and bad conscience. Their responses were dichotomized for this analysis (0 = *not important,* 1 = *important*).[4]

We fitted one- and two-dimensional normal PDF and normal ogive models to these data. We also fitted a restricted two-dimensional model suggested in Jöreskog and Sörbom (1996). In this restricted model Items {1,2,4,5} are taken as indicators of the latent trait *equality*, and Items {3,6,7,8} are taken as indicators of the latent trait *morality*. The two latent traits are correlated. A priori, we do not believe that a model based on a proximity mechanism is suitable for these data. Rather, we expected the normal ogive model with its monotonic item response

[4]Five respondents had out-of-range values for some of the items. Therefore the actual sample size used in the analysis was 195.

TABLE 4
Goodness-of-Fit Tests for the Morality and Equality Data

	T	$\overline{T}$	df	p	$\overline{\overline{T}}$	df	p
Normal PDF							
One trait	0.24	17.48	20	.62	14.22	16.28	.60
Two traits, unrestricted	0.13	12.16	13	.51	10.69	11.43	.51
Two traits, restricted	0.24	17.44	19	.56	14.14	15.40	.54
Normal ogive							
One trait	0.30	16.93	20	.66	12.34	14.57	.62
Two traits, unrestricted	0.11	7.20	13	.89	5.53	9.98	.85
Two traits, restricted	0.26	14.60	19	.75	10.64	13.84	.70

Note. $N = 195$; $T = N\hat{F}$; $\overline{T}$ denotes T adjusted by its asymptotic mean; $\overline{\overline{T}}$ denotes T adjusted by its asymptotic mean and variance.

functions to fit better these data. This is because a priori we expected that the higher the sense of morality and equality of respondents the more likely they would endorse these items.

However, as can be seen in Table 4 all models provide a good fit to the data. Furthermore, the models with two latent traits do not appear to outperform the one-dimensional models. Also, we see in this table that the difference in fit between the one-dimensional normal PDF and normal ogive models is negligible. Yet, the estimated item response functions of these two models are markedly different for most items. This is shown in Figure 4 where we provide plots for selected items.

If we cannot choose between these models based on their fit to the univariate and bivariate moments, could we choose between them using full information statistics? To answer this question, we estimated both unidimensional models using full information maximum likelihood. We obtained $X^2 = 230.68$ and $G^2 = 152.30$ for the normal PDF model and $X^2 = 234.10$ and $G^2 = 148.28$ for the normal ogive model. The number of degrees of freedom is 239. Thus, although a priori we considered that the normal PDF model is not an appropriate model for these data, the fit of the model is only negligibly outperformed by the normal ogive model. A larger data set would be needed to distinguish between these two models in this application.

To shed more light into this issue, we computed EAP scores for both models using the ULS parameter estimates via Equation (12) assuming a standard normal prior distribution for the latent trait. We provide in Figure 5 a scattergram of the EAP scores. As can be seen, for most respondents the EAP scores under both models are approximately linearly related. However, this is not true for a small set of subjects, so that the overall correlation between the EAP scores is only 0.59.

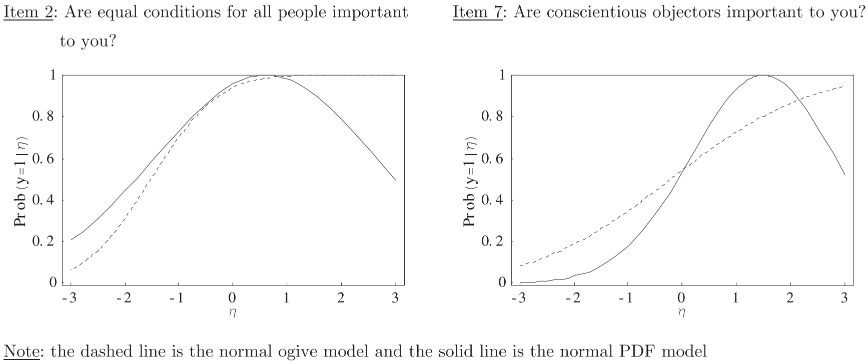


FIGURE 4 Morality and equality data. Plot of selected item response functions.

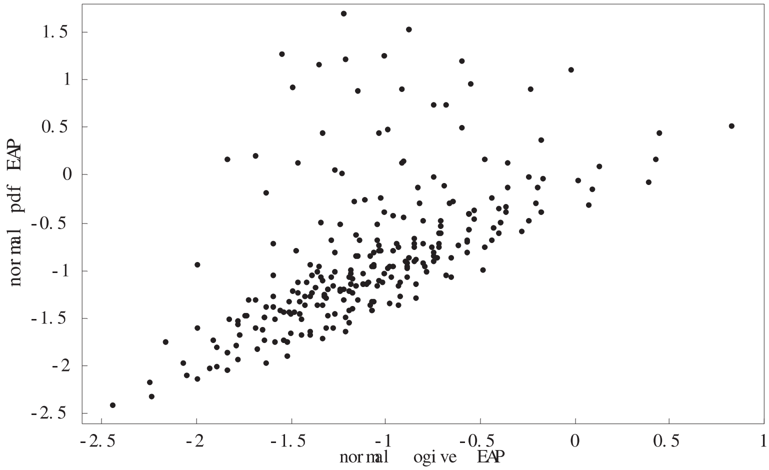


FIGURE 5 Morality and equality data. Scattergram of EAP scores for the normal PDF model against EAP scores for the normal ogive model.

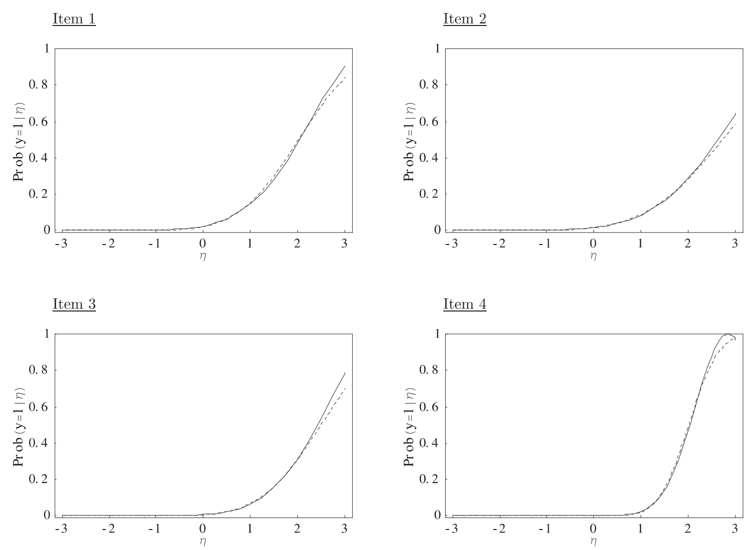
Political Action Survey

Jöreskog and Moustaki (2001) modeled the U.S. sample of the Political Action Survey. The data set consists of six items measuring political efficacy. There were 1,719 individuals who responded to these items using the following categories: *strongly agree, agree, disagree, strongly disagree, do not know,* and *no answer*. After eliminating those cases with *do not know* and *no answer* responses, the total sample size is 1,554. For purposes of illustration, the data were dichotomized as 1 (*strongly disagree*) and 0 (*all else*) for the present analysis.

We estimated one-dimensional normal PDF and normal ogive models to the data dichotomized in this fashion. We obtained the following goodness-of-fit statistics for the normal PDF model: $T = 0.04$, $\bar{T} = 15.39$ on 9 *df*, $p = .08$, and $\bar{\bar{T}} = 13.72$ on 8.03 *df*, $p = .09$. For the normal ogive model estimated using NOHARM we obtained $T = 0.04$, $\bar{T} = 16.19$ on 9 *df*, $p = .06$, and $\bar{\bar{T}} = 11.9$ on 6.62 *df*, $p = .09$. Thus, both models yield a similar fit to the univariate and bivariate moments of these data even though the sample size is large in this case.

This occurs because with this dichotomization the proportion of respondents in Category 1 is very low for all items: {0.08, 0.05, 0.04, 0.02, 0.02}. As a result, the item response functions of the normal PDF model are monotonically increasing in the area of high density of respondents, $\eta \in (-3,3)$, and they are very hard to distinguish from the item response functions of the normal ogive model. This is shown in Figure 6, where we plot the item response functions for the normal PDF and normal ogive models for the first four items.

In fact, the predictions of these two models are hard to distinguish even when all the information available in these data is employed. We estimated by full information maximum likelihood the two competing models. For the normal PDF model we obtained $X^2 = 44.84$ and $G^2 = 52.14$ on 51 *df*. For the normal ogive model we obtained $X^2 = 44.89$ and $G^2 = 53.04$ also on 51 *df*.



Note: the dashed line is the normal ogive model and the solid line is the normal PDF model

FIGURE 6 Political efficacy data. Plot of selected item response functions

CONCLUSIONS

In some item response applications it is reasonable to assume that individuals use a proximity mechanism in responding to the items. Yet, in other applications a cumulative response mechanism seems a priori more reasonable. Latent trait models based on a proximity response mechanism are generally referred to as ideal point models and also as unfolding models. When the data contain moderate amounts of multidimensionality, IRT models with several latent traits may be needed to reproduce the data adequately. To obtain a multidimensional IRT model with nonmonotonic item response functions we simply link a linear function of the latent traits using a normal probability density function. The model proposed is indeed an ideal point model in the sense that a respondent—precisely at the ideal point (the mode of the item response function)—endorses the item with probability one. A more general model with an additional parameter controlling the probability of the modal point is easily conceived. Such a model may deserve future investigation, though problems can be anticipated related to bounds on the probabilities and possibly identifiability. Further research should also consider the extension of this model to the polytomous case.

In the model proposed here the latent trait density is assumed to be multivariate normal. We show that the two specifications of this model (normal probability density link function and normally distributed latent traits) result in closed form expressions for the moments of the multivariate Bernoulli distribution. As a result, cell probabilities under this model can be computed without resorting to numerical integration regardless of the number of traits involved.

We have seen that in applications where a proximity response mechanism is plausible the model indeed fits better than the multidimensional normal ogive model. The model seems particularly suited to model items of the type "I don't have a clear opinion on the topic." Yet, as we have seen, when only a small sample is available the normal PDF and normal ogive models may be hard to distinguish even though their item response functions are quite distinct. Furthermore, even when large samples are available, if the proportion of respondents endorsing the items is very low for all items it may be hard to distinguish these two models as their item response functions will coincide in the region of high density of respondents. Thus, only fitting models based on a proximity mechanism (such as the normal PDF model introduced here) or fitting models based on a cumulative response process (such as the normal ogive model) is to be discouraged. Rather, researchers should compare the fit of competing models based on alternative response processes. Most important, they should consider a priori which response mechanism is most plausible for the application of interest.

To estimate our model we simply minimized the sum of squared errors between the observed and expected univariate and bivariate moments of the MVB distribution, very much as the normal ogive model is estimated in NOHARM.

Limited information methods such as those employed here are an attractive option to estimate latent trait models in sparse binary tables. They are computationally very efficient, they naturally yield limited information test statistics, and they yield residuals which are asymptotically standard normal and that can be readily used to detect the source of misfit in poorly fitting models, as we have illustrated in one of the applications.

Unlike full information test statistics such as Pearson's χ^2 or the likelihood ratio test G^2, limited information test statistics maintain their nominal Type I error rates even in sparse tables. Maydeu-Olivares (2001a) reported a simulation study where 100 observations sufficed to obtain accurate Type I errors using the test statistic employed in this article when testing a six-dimensional IRT model for 21 binary variables. Furthermore, limited information test statistics may be asymptotically more powerful than full information statistics (Maydeu-Olivares & Joe, 2005). Thus, limited information test statistics can be safely used to compare competing models such as the multidimensional PDF model versus the multidimensional normal ogive model.

The model introduced here is not the first multidimensional IRT model with nonmonotonic item response functions. Alternative models have been proposed in the literature by Bradlow and Schmittlein (2000), DeSarbo and Hoffman (1986, 1987), and Takane (1996, 1998). DeSarbo and Hoffman introduced a model that allows representation of both subjects and objects in a joint space and they applied their model to investigate market structure. The respondent latent "disutility" for a certain product involves a weighted distance between the respondent's ideal point and the product (i.e., object) coordinates, an individual threshold value, and an error component with a logistic distribution function. These assumptions lead to a logistic item response function. Estimation is performed via maximum likelihood methods. Similar ideas can be found in Takane (1996, 1998): Respondents and item categories are represented as points in a joint space, and the item response function decreases as the distance between them decreases. After dichotomizing the multiple-choice data, the respondent point is assumed to have a Gaussian distribution and is then integrated out to derive marginal probabilities of response patterns. The item response function is recommended to have a Gaussian form whose parameters depend on the distance. Estimation proceeds by maximum likelihood using the EM algorithm. Finally, Bradlow and Schmittlein (2000) proposed a proximity model to evaluate the performance of six Web search engines to locate web pages. A squared Mahalanobis distance between engines and web pages is defined in a joint space. The probability of an engine finding a given web page is an inverse function (no exponential involved) of that distance. A hierarchical Bayesian model involving inverse Wishart and normal distributions is fitted, and inference is derived using MCMC methods. Goodness-of-fit comparisons are based on the natural log of the Bayes factor. All of these models can be considered as genuine ideal points models while our method, as already stated, becomes an ideal hyperplane model when multidimensional latent variables are considered.

A direction of future research is to compare their approach with ours, and against the normal ogive model (or other multidimensional parametric models based on a cumulative response process) to give additional information on whether proximity or cumulative response mechanisms are more suitable for modeling behavioral, personality, and attitudinal data. Readers interested in this topic should also consult the recent article by Stark, Chernyshenko, Drasgow, and Williams (2006) who, within a unidimensional context, have compared the two-parameter logistic model against a dichotomous version of the generalized graded unfolded model (see Roberts, 2001; Roberts, Donoghue, & Laughlin, 2000; Roberts & Laughlin, 1996) in fitting the scales of the 16PF (Conn & Rieke, 1994).

A final caveat: The normal PDF model is not equivariant to simple operations such as reverse scoring of the items. When estimated with equivariant estimators such as the bivariate ULS estimator employed here or the ML estimator, an equivariant model leads to a probability model in the same parametric family, with parameters transformed in a simple way. For instance, the normal ogive model is equivariant. Reverse scoring all the items leads to a model with the same goodness of fit, and the same parameter estimates, except for a sign change in the intercepts $\boldsymbol{\alpha}$. In contrast, applying the normal PDF model to a dataset after reverse scoring all the items leads to a model with different parameter estimates, and different goodness of fit. To illustrate, we fitted one- and two-dimensional normal PDF models to the satisfaction with life data after reverse scoring all the items. For the one dimensional model we obtained $T = 0.54$, $\bar{T} = 27.82$ on 5 df, $p = .00$, and $\bar{\bar{T}} = 25.07$ on 4.51 df, $p = .00$. For the two-dimensional model we obtained $T = 0.02$, $\bar{T} = 0.90$ on 1 df, $p = .34$, and $\bar{\bar{T}} = 0.90$ on 1 df, $p = .34$. Thus, the normal PDF model fits much better when we model dissatisfaction with life (i.e., after reverse scoring all the items) than when we model satisfaction with life (no reverse scoring).

We conjecture that for an IRT model to be equivariant to reverse scoring, the item response function should be the cumulative density function of a symmetric random variable. Should our conjecture prove correct, then maybe no ideal point model is equivariant. This has important implications for the application of ideal point models to personality research but not so much for their application to attitudinal research. In many attitudinal items there is a natural category to be modeled, so it may not be meaningful to reverse score the items. In contrast, in personality research, often which category is modeled is arbitrary. One instance is modeling satisfaction with life. It is just as meaningful to model dissatisfaction with life. Another instance is modeling extraversion. It is just as meaningful to model introversion. When a model is not equivariant, then it should be fit separately with and without reverse ordering for comparisons, particularly when there is no natural category to be modeled.

ACKNOWLEDGMENTS

Albert Maydeu-Olivares's research was supported by the Department of Universities, Research and Information Society (DURSI) of the Catalan Government and by grants BSO2000-0661 and BSO2003-08507 from the Spanish Ministry of Science and Technology. We are grateful to the referees for comments leading to improvements.

REFERENCES

Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology, 49,* 347–365.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis.* London: Arnold.

Bradlow, E. T., & Schmittlein, D. C. (2000). The little engines that could: Modeling the performance of World Wide Web search engines. *Marketing Science, 19,* 43–62.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–250) New York: Plenum.

Browne, M. W., & du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research, 27,* 269–300.

Chave, E. J., & Thurstone, L. L. (1931). *The measurement of social attitudes.* Chicago: University of Chicago Press.

Chernyshenko, O. S., Stark, S., Chan, K.-Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36,* 563–562.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40,* 5–32.

Conn, S., & Rieke, M. L. (Eds.). (1994). *The 16PF fifth edition technical manual.* Champaign, IL: Institute for Personality and Ability Testing.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

DeSarbo, W. S., & Hoffman, D. L. (1986). Simple and weighted unfolding threshold models for the spatial representation of "pick any/n" data. *Psychometrika, 54,* 105–129.

DeSarbo, W. S., & Hoffman, D. L. (1987). Constructing MDS joint spaces from binary choice data: A multidimensional unfolding threshold model for marketing research. *Journal of Marketing Research, 24,* 40–54.

Diener, E., Emmons, R., Larsen, J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75.

Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research, 23,* 267–269.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research, 21,* 347–387.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8. User's reference guide.* Chicago: Scientific Software.

Levine, M. V. (1984). *An introduction to multilinear formula score theory* (Measurement series 84-4). Champaign, IL: Model Based Measurement Laboratory.

Maydeu-Olivares, A. (2001a). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika, 66,* 209–228.

Maydeu-Olivares, A. (2001b). Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates. *Journal of Educational and Behavioral Statistics, 26,* 49–69.

Maydeu-Olivares, A. (2005). Further empirical results on parametric vs. non-parametric IRT modeling of Likert-type personality data. *Multivariate Behavioral Research, 40,* 275–293.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in 2^n tables: A unified approach. *Journal of the American Statistical Association, 100,* 1009–1020.

McDonald, R. P. (1997). Normal ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257–269). New York: Springer.

McDonald, R. P. (1999). *Test theory. A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Muthén, B. (1993). Goodness of fit with categorical and other non normal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.

Rao, J. N. K., & Scott, A. J. (1987). On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics, 15,* 385–397.

Rosander, A. C., & Thurstone, L. L. (1931). Scale of attitude toward censorship: Scale No. 28. In E.J. Chave & L. L. Thurstone (Eds.), *The measurement of social attitudes*. Chicago: University of Chicago Press.

Roberts, J. S. (1995). Item Response Theory approaches to attitude measurement (Doctoral dissertation, University of South Carolina, Columbia, 1995). *Dissertation Abstracts International, 56,* 7089B.

Roberts, J. S. (2001). GGUM2000: Estimation of parameters in the generalized graded unfolding model. *Applied Psychological Measurement, 25,* 38.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24,* 3–32.

Roberts, J. S., & Laughlin, J. E. (1996). A unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement, 20,* 231–255.

Samejima, F. (1969). Calibration of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement, No. 17.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91,* 25–39.

Takane, Y. (1996). An item response model for multidimensional analysis of multiple-choice data. *Behaviormetrika, 23,* 153–167.

Takane, Y. (1998). Choice model analysis of the “pick any/n” type of binary data. *Japanese Psychological Research, 40,* 31–39.

Teugels, J. L. (1990). Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis, 32,* 256–268.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

van Schuur, W. H., & Kiers, H. A. L. (1994). Why factor analysis is often the incorrect model for analyzing bipolar concepts, and what model can be used instead. *Applied Psychological Measurement, 5,* 245–262.

APPENDIX

To prove (8) we apply the change of variable $\mathbf{z} = \mathbf{\Psi}^{-\frac{1}{2}}\boldsymbol{\eta}$ to obtain $\boldsymbol{\eta} = \mathbf{\Psi}^{\frac{1}{2}}\mathbf{z}$ and $\left|\frac{d\boldsymbol{\eta}}{d\mathbf{z}}\right| = |\mathbf{\Psi}|^{\frac{1}{2}}$. Thus,

$$\Pr\left[\bigcap_{i\in\mathbf{s}}(Y_i=1)\right] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi_p\left(\boldsymbol{\eta}:\mathbf{0},\mathbf{\Psi}\right)\left[\prod_{i\in\mathbf{s}}\sqrt{2\pi}\phi_1\left(\alpha_i+\boldsymbol{\beta}_i'\boldsymbol{\eta}\right)\right]d\boldsymbol{\eta} =$$
$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi_p\left(\mathbf{z}:\mathbf{0},\mathbf{I}\right)\left[\prod_{i\in\mathbf{s}}\sqrt{2\pi}\phi_1\left(\alpha_i+\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\mathbf{z}\right)\right]d\mathbf{z} =$$
$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-\frac{p}{2}}\exp\left\{-\frac{1}{2}\left[\mathbf{z}'\mathbf{z}+\sum_{i\in\mathbf{s}}\left(\alpha_i+\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\mathbf{z}\right)^2\right]\right\}d\mathbf{z}$$

Now, inside the squared brackets we have

$$\mathbf{z}'\mathbf{z}+\sum_{i\in\mathbf{s}}\left(\alpha_i+\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\mathbf{z}\right)^2 = \mathbf{z}'\mathbf{z}+\sum_{i\in\mathbf{s}}\alpha_i^2+2\sum_{i\in\mathbf{s}}\alpha_i\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\mathbf{z}+\sum_{i\in\mathbf{s}}\mathbf{z}'\mathbf{\Psi}^{\frac{1}{2}}\boldsymbol{\beta}_i\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\mathbf{z} =$$
$$= \mathbf{z}'\left(\mathbf{I}+\sum_{i\in\mathbf{s}}\mathbf{\Psi}^{\frac{1}{2}}\boldsymbol{\beta}_i\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\right)\mathbf{z}+2\sum_{i\in\mathbf{s}}\alpha_i\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}}\mathbf{z}+\sum_{i\in\mathbf{s}}\alpha_i^2 =$$
$$= \left(\mathbf{z}-\boldsymbol{\mu}\right)'\mathbf{A}\left(\mathbf{z}-\boldsymbol{\mu}\right)+\mathbf{C}$$

where

$$\mathbf{A} = \mathbf{I}_p+\sum_{i\in\mathbf{s}}\mathbf{\Psi}^{\frac{1}{2}}\boldsymbol{\beta}_i\boldsymbol{\beta}_i'\mathbf{\Psi}^{\frac{1}{2}} = \mathbf{I}+\mathbf{\Psi}^{\frac{1}{2}}\mathbf{B}_s{}'\mathbf{B}_s\mathbf{\Psi}^{\frac{1}{2}},$$

$$\boldsymbol{\mu} = -\mathbf{A}^{-1}\mathbf{\Psi}^{1/2}\left(\sum_{i\in\mathbf{s}}\alpha_i\boldsymbol{\beta}_i\right) = -\mathbf{A}^{-1}\mathbf{\Psi}^{1/2}\mathbf{B}_s{}'\boldsymbol{\alpha}_s,$$

$$\mathbf{C} = \sum_{i\in\mathbf{s}}\alpha_i{}^2-\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\alpha}_s{}'\boldsymbol{\alpha}_s-\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu},$$

and $\boldsymbol{\alpha}_s$ and $\mathbf{B}_s$ are a $k\times 1$ vector and a $k\times p$ matrix obtained by selecting rows in $\boldsymbol{\alpha}$ and $\mathbf{B}$ according to $\mathbf{s}$. Thus, $\Pr\left[\bigcap_{i\in\mathbf{s}}(Y_i=1)\right] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-\frac{p}{2}}\exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}-\boldsymbol{\mu}\right)'\mathbf{A}\right.\right.$

$(\mathbf{z}-\boldsymbol{\mu})+\mathbf{C}\big]\Big\}d\mathbf{z}=\frac{1}{|\mathbf{A}|^{1/2}}\exp\left(-\frac{1}{2}\mathbf{C}\right)$, as $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-\frac{p}{2}}|\mathbf{A}|^{1/2}\exp\Big\{-\frac{1}{2}\big[(\mathbf{z}-\mu)'$
$A(\mathbf{z}-\mu)\big]\Big\}d\mathbf{z}=1.$

Finally, to prove (8) it suffices to show that

$$|\mathbf{A}|=|\boldsymbol{\Sigma}_s| \tag{A.1}$$

$$\mathbf{C}=\boldsymbol{\alpha}_s{}'\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\alpha}_s, \tag{A.2}$$

where $\boldsymbol{\Sigma}_s=\mathbf{I}_k+\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'$.

Now, to prove (A.1) we write $\mathbf{A}=\mathbf{I}_p+\mathbf{H}\mathbf{H}'$ with $\mathbf{H}=\boldsymbol{\Psi}^{\frac{1}{2}}\mathbf{B}_s{}'$. Therefore, $|\mathbf{A}|=|\mathbf{I}_p+\mathbf{H}\mathbf{H}'|=|\mathbf{I}_k+\mathbf{H}'\mathbf{H}|=\left|\mathbf{I}_k+\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\right|=|\boldsymbol{\Sigma}_g|$.

To prove (A.2), as $\mathbf{A}^{-1}=\mathbf{I}_p-\mathbf{H}\left(\mathbf{I}_k+\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'=\mathbf{I}_p-\boldsymbol{\Psi}^{\frac{1}{2}}\mathbf{B}_s{}'\boldsymbol{\Sigma}_s^{-1}\mathbf{B}_s\boldsymbol{\Psi}^{\frac{1}{2}}$ and $\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'=\boldsymbol{\Sigma}_s-\mathbf{I}_k$,

$$\begin{aligned}
\mathbf{C}&=\boldsymbol{\alpha}_s{}'\boldsymbol{\alpha}_s-\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}=\boldsymbol{\alpha}_s{}'\boldsymbol{\alpha}_s-\boldsymbol{\alpha}_s{}'\mathbf{B}_s\boldsymbol{\Psi}^{\frac{1}{2}}\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\boldsymbol{\Psi}^{\frac{1}{2}}\mathbf{B}_s{}'\boldsymbol{\alpha}_s=\\
&=\boldsymbol{\alpha}_s{}'\left(\mathbf{I}_k-\mathbf{B}_s\boldsymbol{\Psi}^{\frac{1}{2}}\mathbf{A}^{-1}\boldsymbol{\Psi}^{\frac{1}{2}}\mathbf{B}_s{}'\right)\boldsymbol{\alpha}_s=\boldsymbol{\alpha}_s{}'\left(\mathbf{I}_k-\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'+\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\boldsymbol{\Sigma}_s^{-1}\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\right)\boldsymbol{\alpha}_s=\\
&=\boldsymbol{\alpha}_s{}'\left[\mathbf{I}_k+\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\left(-\mathbf{I}_k+\boldsymbol{\Sigma}_s^{-1}\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\right)\right]\boldsymbol{\alpha}_s=\\
&=\boldsymbol{\alpha}_s{}'\left[\mathbf{I}_k+\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\left(-\mathbf{I}_k+\mathbf{I}_k-\boldsymbol{\Sigma}_s^{-1}\right)\right]\boldsymbol{\alpha}_s=\\
&=\boldsymbol{\alpha}_s{}'\left(\mathbf{I}_s-\mathbf{B}_s\boldsymbol{\Psi}\mathbf{B}_s{}'\boldsymbol{\Sigma}_s^{-1}\right)\boldsymbol{\alpha}_s=\boldsymbol{\alpha}_s{}'\left(\mathbf{I}_k-\mathbf{I}_k+\boldsymbol{\Sigma}_s^{-1}\right)\boldsymbol{\alpha}_s=\boldsymbol{\alpha}_s{}'\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\alpha}_s
\end{aligned}$$

which concludes the proof.