P1: FCH/FFX P2: FCH/FFX QC: FCH/FFX T1: FCH
LE104-03 LE104-Maydeu-v2.cls February 20, 2005 13:49

3

Linear Item Response Theory, Nonlinear Item Response Theory, and Factor Analysis: A Unified Framework

Albert Maydeu-Olivares
University of Barcelona

INTRODUCTION

What it is now known as item response modeling [for an overview see van der Linden and Hambleton (1997)] originated as an effort to overcome the limitations of the factor model when applied to test items. Test items are most often categorical in nature, whereas the factor model was designed for continuous data. Unfortunately, over the years item response modeling and factor modeling have developed rather independently from one another. One of the recurring topics in R. P. McDonald's career has been establishing bridges between these two fields (McDonald, 1967, 1980, 1981, 1982a, 1982b, 1985a, 1985b, 1986, 1999, 2001; McDonald & Mok, 1995). Two approaches can be used to relate the nonlinear models used in item response theory (IRT) to the linear model used in factor analysis. One approach is to use harmonic analysis (e.g., McDonald, 1967, 1982a). The second approach is to use link functions (e.g., McDonald, 1999; Moustaki & Knott, 2000).

This chapter focuses on one particular item response model for binary data, the linear IRT model. In this model, the conditional probability of endorsing an item given the latent traits is simply a linear function. McDonald (1999, chap. 12 and 13; see also McDonald, 1969, 1982a) discussed at length the application of the usual IRT theoretical machinery (e.g., information functions) to this model. McDonald (1999) also pointed out that when this model is estimated using bivariate information, it is equivalent to the factor model. In this chapter we explore further the linear IRT model for binary data and its relation to the factor model. We show that with binary data these two models are not always equivalent. In fact, they are only equivalent when the linear IRT model is estimated using only univariate and bivariate information. Thus, in relating the factor model to the linear item response model it is necessary to take into account estimation issues, in particular the use of limited- versus full-information methods. The use of limited- versus full-information estimation methods in IRT is discussed by Bolt (chap. 2, this volume; see also Maydeu-Olivares, 1996), and Krane and Slaney (chap. 5, this volume) provide an useful introduction to the factor model; a more detailed presentation of IRT modeling is given by Ackerman (chap. 1, this volume).

This chapter is organized as follows. In the next section we discuss the linear item response model within a general presentation of item response models using link functions. The use of harmonic analysis as a unifying framework for both linear and nonlinear item response models is discussed at the end of the section. The third section discusses the factor model and its application to binary data. In that section we relate the factor model to the linear item response model. The fourth section is devoted to estimation and testing. First, we discuss estimation and testing in factor analysis. Next, we discuss estimation and testing in IRT. We close that section by describing some of the challenges currently faced in estimating and testing IRT models and introduce new theoretical results that address these challenges. Several numerical examples are provided in the final section to illustrate the discussion.

THE LINEAR ITEM RESPONSE MODEL FOR BINARY DATA

Item Response Modeling for Binary Data: Nonlinear Models

Consider n binary variables $\mathbf{y} = (y_1, \ldots, y_n)'$, each one with two possible outcomes. Without loss of generality, we may assign the values $\{0, 1\}$ to these possible outcomes. Therefore, the distribution of each y_i is Bernoulli, and the joint distribution of $\boldsymbol{y}$ is multivariate Bernoulli (MVB).

All item response models for binary data take on the form (e.g., Bartholomew and Knott, 1999)

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

where $\Pr(\bigcap_{\mathrm{i}=1}^{n} y_i)$ denotes the probability of observing one of the possible 2^n binary patterns, $\gamma_p(\boldsymbol{\eta})$ denotes the probability density function of a p-dimensional vector of continuous *unobserved* latent traits $\boldsymbol{\eta}$, and $\Pr(y_i = 1|\boldsymbol{\eta})$ is usually denoted as the *item response function* (IRF).

Let $z_i = \alpha_{\mathrm{i}} + \boldsymbol{\beta}_i'\boldsymbol{\eta}$, where α_i is an intercept and $\boldsymbol{\beta}_i$ is a $p \times 1$ vector of slopes. Two widely used IRFs are

$$\Pr(y_i = 1|\boldsymbol{\eta}) = \Phi_1(z_i) = \int_{-\infty}^{\alpha_i+\beta_i'\eta} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt, \tag{2}$$

$$\Pr(y_i = 1|\boldsymbol{\eta}) = \Psi(z_i) = \frac{1}{1 + e^{-(\alpha_i+\beta_i'\eta)}}, \tag{3}$$

where $\Phi_1(z_i)$ and $\Psi(z_i)$ denote, respectively, univariate standard normal and standard logistic distribution functions evaluated at z_i. These functions link z_i to the probability of endorsing an item, given a fixed value of the latent traits.

Now, to completely specify Equation 1 we also need to specify the density of the latent traits, $\gamma_p(\boldsymbol{\eta})$. This is generally assumed to be multivariate normal with mean zero and some correlation matrix $\boldsymbol{\Phi}$, that is,

$$\gamma_p(\boldsymbol{\eta}) = \phi_p(\boldsymbol{\eta} : \mathbf{0}, \boldsymbol{\Phi}). \tag{4}$$

The model given by Equation 1 with Equations 2 and 4 is referred to as the *multidimensional normal ogive model*, whereas the model given by Equation 1 with Equations 3 and 4 is referred to as the *multidimensional two-parameter logistic model*. Note, however, that the IRFs given by Equations 2 and 3 can be coupled in fact with any density function $\gamma_p(\boldsymbol{\eta})$, for instance, with a nonparametric function. Similarly, the IRF can also be a nonparametric function.

Generally, we require two properties from an IRF:

Property 1. An IRF should be bounded between 0 and 1 because it is a probability.

Property 2. An IRF should be smooth.

In addition, when modeling cognitive test items, we generally also require the following:

Property 3. An IRF should be monotonically increasing.

In the case of attitudinal or personality items, it has been argued (e.g., van Schuur & Kiers, 1994) that Property 3 need not be a reasonable assumption. The IRFs given by Equations 2 and 3 are monotonically increasing. A non-monotonically increasing multidimensional IRF is

$$\Pr(y_i = 1|\boldsymbol{\eta}) = \sqrt{2\pi}\ \phi_1(z_i) = e^{-(\alpha_i+\beta_i'\eta)^2/2}, \tag{5}$$

where $\phi_1(z_i)$ denotes a univariate standard normal density function evaluated at z_i. Maydeu-Olivares, Hernández, and McDonald (2004) recently introduced a model with the IRF given by Equation 5 and normally distributed latent traits, which they denote the *normal PDF model*. The normal ogive, the two-parameter logistic, and the normal PDF models are obtained by simply using the nonlinear functions $\Phi_1(z_i)$, $\Psi(z_i)$, and $\sqrt{2\pi}\phi_1(z_i)$ to link z_i to $\Pr(y_i = 1|\boldsymbol{\eta})$.

The Linear Item Response Model

The linear item response model for binary data discussed in McDonald (1999) simply amounts to using an identity link function $I(z_i)$ instead of a nonlinear link function to specify the IRF. Thus, the IRF of this model is

$$\Pr(y_i = 1|\boldsymbol{\eta}) = I(z_i) = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i'\boldsymbol{\eta}. \tag{6}$$

The IRF of this model violates Property 1 because it is not bounded between 0 and 1. Thus, for large enough values of the latent traits it yields probabilities larger than 1, and for small enough values it yields probabilities less than 0 (McDonald, 1999). This is a very unappealing property of the model.

On the other hand, the linear model enjoys a very attractive property that has not been noticed, namely, we need not specify a latent trait density. This can be readily seen if we characterize the multivariate Bernoulli distribution using its joint raw moments. In the Appendix we discuss two alternative representations of this distribution: (a) using the set of 2^n binary pattern probabilities $\boldsymbol{\pi}$ and (b) using the set of $2^n - 1$ joint raw moments of this distribution $\dot{\boldsymbol{\pi}}$. We also show that there is a one-to-one relationship between these two representations.

Consider, for example, a unidimensional linear latent trait model for $n = 3$ items. Let κ_i denote the ith raw moment of the latent trait,

$$\kappa_i = E[\eta^i], \tag{7}$$

so that, for instance, the mean of the latent trait is denoted by κ_1. Notice that there are n latent trait moments in a unidimensional linear IRT model for n variables. Using Equations 6 and 7, with Equation A7 of the Appendix, we obtain the univariate moments of the MVB distribution under the linear IRT model as

$$\dot{\pi}_i = \mathrm{E}_\eta[\alpha_i + \beta_i\eta] = \alpha_i + \beta_i \mathrm{E}[\eta] = \alpha_i + \beta_i\kappa_1. \tag{8}$$

Similarly, using Equation A8, we obtain the bivariate raw moments of the MVB distribution under this model as

$$\begin{aligned}\dot{\pi}_{ij} &= \mathrm{E}_\eta[(\alpha_i + \beta_i\eta)(\alpha_j + \beta_j\eta)] = \alpha_i\alpha_j + (\alpha_i\beta_j + \alpha_j\beta_i)E[\eta] + \beta_i\beta_j E[\eta^2] \\ &= \alpha_i\alpha_j + (\alpha_i\beta_j + \alpha_j\beta_i)\kappa_1 + \beta_i\beta_j\kappa_2.\end{aligned} \tag{9}$$

P1: FCH/FFX P2: FCH/FFX QC: FCH/FFX T1: FCH
LE104-03 LE104-Maydeu-v2.cls February 20, 2005 13:49

Finally, using Equation A9, we obtain the trivariate moments under this model as

$$\dot{\pi}_{ijk} = \alpha_i\alpha_j\alpha_k + (\alpha_i\alpha_j\beta_k + \alpha_i\alpha_k\beta_j + \alpha_j\alpha_k\beta_i)\kappa_1 \\ + (\alpha_i\beta_j\beta_k + \alpha_j\beta_i\beta_k + \alpha_k\beta_i\beta_j)\kappa_2 + \beta_i\beta_j\beta_k\kappa_3. \quad (10)$$

This example illustrates how the moments of **y** under the linear IRT model only depend on the item parameters and on the moments of the latent traits, regardless of the density of the latent traits. This is also true of the cell probabilities because of Equation A5 (see the Appendix).

Not all parameters of the linear IRT model are identified. Fixing any two moments to 0 and 1, respectively, suffices to identify a unidimensional model. These two fixed moments set the location and scale of the latent trait. This can be checked by verifying that $\boldsymbol{\Delta} = \partial\boldsymbol{\pi}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ is of full rank (Bekker, Merckens & Wansbeek, 1994), where $\boldsymbol{\theta}$ denotes the model parameters (i.e., the item parameters and the moments of the latent traits) stacked in a column vector. Thus, if, for instance, (a) the mean and the variance of the latent trait are fixed to 0 and 1, respectively, or (b) the nth- and (nth − 1) order moments of the latent trait are fixed to 0 and 1, respectively, then all the item parameters and the remaining moments of the latent trait are identified.

Fewer parameters can be identified when the model is estimated using limited-information methods. For instance, suppose that the model is to be estimated using only univariate and bivariate information. Then third- and higher order moments of the latent traits cannot be identified because they do not appear in Equations 8 and 9. The means and variances of the latent traits cannot be identified either. In this case, the means can be set to 0 and the variances to 1 to identify the model.

In closing our treatment of the linear IRT model, we consider making statements about an individual's location on the latent traits given the individual's 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),

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

Thus, after the item parameters and latent trait moments 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 as maximum a posteriori (MAP) scores. Obtaining MAP scores in general requires an iterative procedure, whereas obtaining EAP scores involves computing

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

It is of interest that in the linear IRT model, although it is not necessary to assume any prior distribution for $\gamma_p(\boldsymbol{\eta})$ to estimate the model parameters, it is necessary to assume some prior distribution to obtain these scores.

Wide-Sense Linear Item Response Models

So far, we have considered linear and nonlinear link functions for obtaining the IRF in item response models. In passing, we point out that a third alternative for obtaining an IRF is to use a wide-sense linear function in the latent traits (McDonald, 1967, 1982a). This is a function that is linear in the item parameters, but nonlinear in the latent traits,

$$\Pr(y_i = 1|\boldsymbol{\eta}) = \alpha_i + \sum_{j=1}^{p} \beta_{ij} \varphi_j (\boldsymbol{\eta}), \tag{13}$$

for some nonlinear functions $\varphi_j(\boldsymbol{\eta})$. A typical example of a wide-sense model is the unidimensional cubic model

$$\Pr(y_i = 1|\eta) = \alpha_i + \beta_{i1}\eta + \beta_{i2}\eta^2 + \beta_{i3}\eta^3. \tag{14}$$

McDonald (1982a) pointed out that wide sense linear models may offer a unified framework for IRFs that encompasses both the linear and nonlinear models as special cases.

It remains to be investigated whether any item response model can be written as a wide-sense model. However, it is easy to show using Hermite polynomials that any item response model with differentiable item response functions and normally distributed latent traits can be expressed as a wide-sense linear model. A Hermite polynomial of degree k, $H_k(x)$, satisfies by definition $H_k(x)\phi(x) = (-1)^k \partial^k \phi(x)/\partial x^k$. The first four terms of this polynomial are

$$H_k(x) = \begin{cases} 1 & \text{if} \quad k = 0, \\ x & \text{if} \quad k = 1, \\ x^2 - 1 & \text{if} \quad k = 2, \\ x^3 - 3x & \text{if} \quad k = 3. \end{cases} \tag{15}$$

For instance, McDonald (1967; see also McDonald, 1997) showed that the unidimensional version of the normal ogive model given by Equation 2 can be written as

$$\Phi_1(\alpha_i + \beta_i \eta) = \Phi_1\left(\frac{\alpha_i}{\sqrt{1+\beta_i^2}}\right) + \phi_1\left(\frac{-\alpha_i}{\sqrt{1+\beta_i^2}}\right) \times \sum_{k=1}^{\infty} \frac{1}{k!}\left(\frac{\beta_i}{\sqrt{1+\beta_i^2}}\right)^k H_{k-1}\left(\frac{-\alpha_i}{\sqrt{1+\beta_i^2}}\right) H_k(\eta). \quad (16)$$

Also, it can be shown that a unidimensional normal PDF model as in Equation 5 can be written as

$$\sqrt{2\pi}\phi_1(\alpha_i + \beta_i \eta) = \sqrt{2\pi}\left(1+\beta_i^2\right)^{-1/2} \phi_1\left(\frac{-\alpha_i}{\sqrt{1+\beta_i^2}}\right) \times \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\beta_i}{\sqrt{1+\beta_i^2}}\right)^k H_k\left(\frac{-\alpha_i}{\sqrt{1+\beta_i^2}}\right) H_k(\eta). \quad (17)$$

A expression for the two-parameter logistic model with a normally distributed latent trait can similarly be obtained, but this does not seem to have been attempted. In any case, we see in Equations 16 and 17 that strictly nonlinear models can be expressed as wide-sense linear models with an infinite number of terms. In practice, they can be well approximated with a small number of terms. For instance, both the normal ogive and normal PDF model can be reasonably approximated by truncating the series in Equations 16 and 17 at $k = 3$.

THE FACTOR MODEL

Description of the Model

Let $\mathbf{y}$ be a $n \times 1$ vector of observed variables to be modeled, $\boldsymbol{\eta}$ be a $p \times 1$ vector of unobserved latent traits (factors), where $n > p$, and $\boldsymbol{\varepsilon}$ be an $n \times 1$ vector of random errors. The factor model assumes that

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (18)$$

where $\boldsymbol{\alpha}$ is an $n \times 1$ vector of intercepts and $\mathbf{B}$ is an $n \times p$ matrix of slopes (factor loadings). The model further assumes that the mean of the latent traits is zero, that the mean of the random errors is zero, and that the latent traits and random errors are uncorrelated. That is,

$$\mathrm{E}\,[\boldsymbol{\eta}] = \mathbf{0},\ \mathrm{E}\,[\boldsymbol{\varepsilon}] = \mathbf{0},\ \mathrm{cov}\,[\boldsymbol{\eta}] = \boldsymbol{\Phi},\ \mathrm{cov}\,[\boldsymbol{\varepsilon}] = \boldsymbol{\Psi},\ \mathrm{cov}[\boldsymbol{\eta}, \boldsymbol{\varepsilon}'] = \mathbf{0}. \tag{19}$$

Furthermore, the random errors are generally assumed to be mutually uncorrelated, so that $\boldsymbol{\Psi}$ is a diagonal matrix.

We note two interesting features about this model: First, no assumptions are made on the distribution of the latent traits $\boldsymbol{\eta}$ nor the errors $\boldsymbol{\varepsilon}$. As a result, no assumptions are made on the distribution of the observed variables $\boldsymbol{y}$.

The factor model as defined by Equations 18 and 19 has an interesting second feature: It is a partially specified model. By this, we mean the following: Under assumptions 18 and 19, it follows that

$$\boldsymbol{\mu} = \mathrm{E}[\mathbf{y}] = \boldsymbol{\alpha} \tag{20}$$

$$\boldsymbol{\Sigma} = \mathrm{cov}[\mathbf{y}\mathbf{y}'] = \mathbf{B}\boldsymbol{\Phi}\mathbf{B}' + \boldsymbol{\Psi}, \tag{21}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the population univariate and bivariate central moments of $\mathbf{y}$, respectively, which depend solely on the model parameters. Moreover, it is a partially specified model in the sense that using only assumptions 18 and 19, we have that the trivariate moments of the observed variables do not depend solely on the model parameters. They also depend, for instance, on the third-order moments of the latent traits. However, these are left unspecified in the factor model. In contrast, item response models are completely specified models in the sense that all the moments of $\mathbf{y}$ are completely specified by the model parameters.

It is not surprising that the factor model is a partially specified model. The objective of factor analysis applications is to model the bivariate associations present in the data: either the central moments (sample covariances) or the standardized central moments (sample correlations). Generally, the mean structure is of no interest, and only the parameters involved in the covariance structure are estimated.

In closing this section, it is interesting that historically it has been frequently assumed that the latent traits $\boldsymbol{\eta}$ and the errors $\boldsymbol{\varepsilon}$ are jointly multinormally distributed. Under this additional assumption (which we do not make here), the distribution of $\mathbf{y}$ is multivariate normal and the factor model becomes a completely specified model because the multivariate distribution is completely specified by its first two moments.

Relationship Between the Factor Model and the Linear Item Response Model

The linear item response model presented here is a model for binary data. In contrast, the factor model does not make any assumptions about the nature of the observed variables. Thus, in principle, it can be applied to binary data. However, when the observed variables are binary, $\mathrm{var}(y_i) = \mu_i(1 - \mu_i)$. Coupling this with Equation 20, we see that when the factor model is applied to binary data it must satisfy

$$\mathrm{var}(y_i) = \alpha_i - \alpha_i^2. \tag{22}$$

where α_i denotes the ith element of $\boldsymbol{\alpha}$. However, from Equation 21 it must also satisfy

$$\mathrm{var}(y_i) = \boldsymbol{\beta}_i'\boldsymbol{\Phi}\boldsymbol{\beta}_i + \psi_i^2. \tag{23}$$

where ψ_i^2 denotes the ith element on the diagonal of $\boldsymbol{\Psi}$, and $\boldsymbol{\beta}_i'$ denotes the ith row of $\mathbf{B}$. As a result, when the factor model is applied to binary data the elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$ are jointly underidentified. In other words, when the factor model is applied to binary data one can estimate either $\{\boldsymbol{\alpha}$, $\mathbf{B}$ and $\boldsymbol{\Phi}\}$ or $\{\boldsymbol{\Psi}$, $\mathbf{B}$ and $\boldsymbol{\Phi}\}$. These are two alternative parametrizations of the model. We refer to the former as the $\boldsymbol{\alpha}$ parametrization and the latter as the $\boldsymbol{\Psi}$ parametrization. Using Equation 22 and 23, we obtain the relationship between these parametrizations as

$$\alpha_i = \frac{1 + \sqrt{1 - 4\boldsymbol{\beta}_i'\boldsymbol{\Phi}\boldsymbol{\beta}_i - 4\psi_i^2}}{2}, \tag{24}$$

$$\psi_i^2 = \alpha_i - \alpha_i^2 - \boldsymbol{\beta}_i'\boldsymbol{\Phi}\boldsymbol{\beta}_i. \tag{25}$$

Note that if the factor model is estimated using only the covariance structure (ignoring the mean structure), this identification problem goes unnoticed because $\boldsymbol{\alpha}$ is not involved. Also notice that in estimating a factor model from binary data all the identified model parameters can be estimated using only the covariance matrix. In this case, it seems natural to use the $\boldsymbol{\Psi}$ parametrization, but one can also use the $\boldsymbol{\alpha}$ parametrization. The covariance structure implied by the α parametrization is, from Equation 25,

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Phi}\mathbf{B}' + \mathrm{diag}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^2) - \mathrm{Diag}(\mathbf{B}\boldsymbol{\Phi}\mathbf{B}'), \tag{26}$$

where we use $\mathrm{diag}(\mathbf{x})$ to indicate a diagonal matrix with diagonal elements equal to $\mathbf{x}$ and $\mathrm{Diag}(\mathbf{X})$ to indicate a matrix where all the off diagonal elements of $\mathbf{X}$ have been set to zero.

Equations 20 and 26 are also the mean and covariance structures implied by the linear item response model. Thus, the factor model applied to binary data and the linear item response model estimated from univariate and bivariate information are equivalent models. In general, they are not equivalent models because the linear item response model can be estimated using full information, and in this case some of the moments of the latent traits can be estimated.

Because the linear item response model and the factor model are equivalent when estimated from bivariate information, a question immediately arises. Can we compare the fit of a factor model (estimated by bivariate methods) and of a nonlinear item response model to a given binary data set? In order to answer this question it is necessary to discuss not only statistical theory for goodness of fit testing but also for estimation in both item response modeling and in factor analysis.

ESTIMATION AND TESTING

Factor Model

Let $\boldsymbol{\theta}$ be the q-dimensional vector of parameters to be estimated. Also, let $\boldsymbol{\sigma}$ be the $t = n(n+1)/2$ -dimensional vector obtained by stacking the elements on the diagonal or below the diagonal of $\boldsymbol{\Sigma}$. Finally, let $\mathbf{s}$ be the sample counterparts of $\boldsymbol{\sigma}$ (i.e., sample variances and covariances). A popular approach to estimate the parameters of the factor model is to minimize the weighted least squares (WLS) function,

$$F = (\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))' \,\hat{\mathbf{W}}\, (\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})), \tag{27}$$

where $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a positive-definite matrix. Now, let $\boldsymbol{\Delta} = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\mathbf{H} = (\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}$. Also, let $\stackrel{d}{\to}$ denote convergence in distribution.

Because $\sqrt{N}(\boldsymbol{s} - \boldsymbol{\sigma}) \stackrel{d}{\to} N(\mathbf{0}, \boldsymbol{\Gamma})$, then, if $\boldsymbol{\Delta}$ is of full rank q and some other mild regularity conditions are satisfied (Browne, 1984), the parameter estimates $\hat{\boldsymbol{\theta}}$ obtained by minimizing Equation 27 are consistent, and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{d}{\to} N(\mathbf{0}, \mathbf{H}\boldsymbol{\Gamma}\mathbf{H}'), \tag{28}$$

$$\sqrt{N}(\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})) \stackrel{d}{\to} N(\mathbf{0}, \boldsymbol{\Omega}), \qquad \boldsymbol{\Omega} = (\mathbf{I} - \boldsymbol{\Delta}\mathbf{H})\boldsymbol{\Gamma}(\mathbf{I} - \boldsymbol{\Delta}\mathbf{H})', \tag{29}$$

where $(\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}))$ denotes the residual variances and covariances.

Some obvious choices of $\hat{\mathbf{W}}$ in Equation 27 are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Gamma}}^{-1}$ (minimum variance WLS, or MVWLS), $\hat{\mathbf{W}} = (\mathrm{Diag}(\hat{\boldsymbol{\Gamma}}))^{-1}$ (diagonally WLS, or DWLS) and $\hat{\mathbf{W}} = \mathbf{I}$ (unweighted least squares, or ULS).

Following Browne (1984), when the factor model is estimated by minimizing Equation 27, we can obtain a goodness-of-fit test of the restrictions imposed by the model on the means and covariances of $\mathbf{y}$ by using

$$T_B = N(\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}))'\hat{\mathbf{U}}(\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})), \qquad \mathbf{U} = \boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-1}. \tag{30}$$

T_B is asymptotically distributed as a chi-square distribution with $t - q$ degrees of freedom regardless of the weight matrix used in Equation 27. To obtain standard errors for the parameter estimates and residuals and to obtain an overall goodness-of-fit test using Equations 28,30, we evaluate $\boldsymbol{\Delta}$ at the parameter estimates and consistently estimate $\boldsymbol{\Gamma}$ using sample third- and fourth-order central moments.

Previously, we referred to the estimator obtained by using $\hat{\mathbf{W}} = \hat{\boldsymbol{\Gamma}}^{-1}$ in Equation 27 as the minimum variance WLS estimator. This is because with this choice of weight matrix, the resulting estimator has minimum variance (asymptotically) within the class of estimators based on the sample covariances. In the case of the MVWLS estimator, Equation 28–30 simplify to

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, (\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1}) \tag{31}$$

$$\sqrt{N}(\boldsymbol{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}(\boldsymbol{\Delta}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'. \tag{32}$$

$$T_B = N\hat{F} \xrightarrow{d} \chi^2_{t-q}. \tag{33}$$

Here, we have focused on the weighted least squares discrepancy function 27 (also denoted as the generalized least squares function). Another discrepancy function that is often used to estimate the factor model is the discrepancy function

$$F = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\mathbf{S}| + \mathrm{tr}((\boldsymbol{\Sigma}(\boldsymbol{\theta}))^{-1}\mathbf{S}) - n, \tag{34}$$

where $\mathbf{S}$ is the sample covariance matrix of $\mathbf{y}$. If $\mathbf{y}$ is normally distributed, minimizing Equation 34 yields maximum likelihood estimates. When $\mathbf{y}$ is not normally distributed, standard errors for the model parameters estimated by minimizing Equation 34 and goodness-of-fit tests can be obtained using Equations 28 and 30, respectively (e.g., Satorra & Bentler, 1994). Another method widely used to assess the goodness of fit when Equation 34 is minimized without a normality assumption and when Equation 27 is minimized using $\hat{\boldsymbol{W}} \neq \hat{\boldsymbol{\Gamma}}^{-1}$ is to adjust $N\hat{F}$ by its mean (or by its mean and variance) so that the resulting test statistic asymptotically matches the mean (or the mean and the variance) of a chi-square distribution with $t - q$ degrees of freedom (Satorra & Bentler, 1994).

Item Response Models

Let $\boldsymbol{\pi}(\boldsymbol{\theta})$ denote the 2^n vector of the binary pattern probabilities of Equation 1 expressed as a function of the q mathematically independent parameters $\boldsymbol{\theta}$ of an item response model, and let $\mathbf{p}$ be the sample counterpart of $\boldsymbol{\pi}$ (i.e., cell proportions). Item response models for binary data are commonly estimated by maximizing the log-likelihood function

$$\ln L = N\,\mathbf{p}'\ln(\boldsymbol{\pi}(\boldsymbol{\theta})). \tag{35}$$

Thus, the resulting parameter estimates $\hat{\boldsymbol{\theta}}$ are maximum likelihood estimates. Instead of maximizing Equation 35, it is convenient to minimize

$$F_{ML} = \mathbf{p}'\ln\left(\frac{\mathbf{p}}{\boldsymbol{\pi}(\boldsymbol{\theta})}\right), \tag{36}$$

Now, let $\mathbf{D} = \text{diag}(\boldsymbol{\pi})$. Since

$$\sqrt{N}\,(\mathbf{p} - \boldsymbol{\pi}) \xrightarrow{d} N\,(\mathbf{0}, \boldsymbol{\Gamma})\,, \quad \boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\pi}\boldsymbol{\pi}', \tag{37}$$

then, if $\boldsymbol{\Delta} = \boldsymbol{\partial\pi}\,(\boldsymbol{\theta})/\boldsymbol{\partial\theta}'$ is of full rank q and some other regularity conditions are satisfied (Agresti, 1990; Rao, 1973), the maximum likelihood parameter estimates are consistent, they have minimum variance (asymptotically), and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, (\boldsymbol{\Delta}'\boldsymbol{D}^{-1}\boldsymbol{\Delta})^{-1}). \tag{38}$$

Also, we have the following result for the residual cell proportions $(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}))$

$$\sqrt{N}(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) \xrightarrow{d} N\,(\mathbf{0}, \boldsymbol{\Omega})\,, \qquad \boldsymbol{\Omega} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'. \tag{39}$$

To obtain standard errors for the parameter estimates and residuals, all matrices are evaluated at the parameter estimates.

The two most widely used statistics to assess the goodness of fit of the model are the likelihood ratio test statistic G^2 and Pearson's X^2 statistic,

$$G^2 = 2N\ \mathbf{p}' \ln\left(\frac{\mathbf{p}}{\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})}\right) = 2N\hat{F}_{ML}, \tag{40}$$

$$X^2 = N(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}))'(\text{diag}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})))^{-1}(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})). \tag{41}$$

When the model holds, both statistics are asymptotically equivalent and they are asymptotically chi-square distributed with $2^n - q - 1$ degrees of freedom.

We now consider an alternative approach to estimating the IRT parameters that is related to the weighted least squares function in Equation 27 used to estimate the factor model and also to Pearson's X^2 statistic. Suppose $\hat{\boldsymbol{\theta}}$ is obtained by minimizing the generalized minimum chi-square function

$$F = (\boldsymbol{p} - \boldsymbol{\pi}\,(\boldsymbol{\theta}))'\,\hat{\mathbf{W}}\,(\boldsymbol{p} - \boldsymbol{\pi}\,(\boldsymbol{\theta}))\,, \tag{42}$$

where $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a positive-definite matrix. Then, if $\boldsymbol{\Delta} = \partial\boldsymbol{\pi}\,(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ is of full rank q and some other regularity conditions are satisfied (Ferguson, 1996), $\hat{\boldsymbol{\theta}}$ is consistent, and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{H}\boldsymbol{\Gamma}\mathbf{H}'), \qquad \mathbf{H} = (\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W} \tag{43}$$

$$\sqrt{N}(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) \xrightarrow{d} N\,(\mathbf{0}, \boldsymbol{\Omega})\,, \qquad \boldsymbol{\Omega} = (\mathbf{I} - \boldsymbol{\Delta}\mathbf{H})\,\boldsymbol{\Gamma}\,(\mathbf{I} - \boldsymbol{\Delta}\mathbf{H})'\,, \tag{44}$$

where $\boldsymbol{\Gamma}$ is given by Equation (37). To obtain the standard error for the parameter estimates and residuals, $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$ are evaluated at the parameter estimates. Some obvious choices of $\hat{\mathbf{W}}$ in Equation 42 are $\hat{\mathbf{W}} = \hat{\mathbf{D}}^{-1}$ and $\hat{\mathbf{W}} = \mathbf{I}$. When $\hat{\mathbf{W}} = \hat{\mathbf{D}}^{-1}$,

P1: FCH/FFX P2: FCH/FFX QC: FCH/FFX T1: FCH
LE104-03 LE104-Maydeu-v2.cls February 20, 2005 13:49

we obtain asymptotically minimum variance estimators and Equations 43 and 44 reduce to Equations 38 and 39, respectively. Note that in this case we can use $\hat{\mathbf{D}}^{-1} = (\text{diag}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})))^{-1}$ or $\hat{\mathbf{D}}^{-1} = (\text{diag}(\mathbf{p}))^{-1}$. The former is the usual minimum chi-square estimator, whereas the latter is the modified minimum chi-square estimator. The two are asymptotically equivalent. When Equation 42 is minimized with $\hat{\mathbf{W}} = (\text{diag}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})))^{-1}$, $N\hat{F} = X^2$.

New Developments in IRT Estimation and Testing

Because statistical inference for item response models generally involves full-information procedures, whereas statistical inference for the factor model involves limited-information procedures, the former are generally computationally more involved than the latter. Furthermore, statistical inference for item response models faces several challenges (Bartholomew & Knott, 1999; Bartholomew & Leung, 2001; Bartholomew & Tzamourani, 1999; Reiser, 1996; Reiser & VandenBerg, 1994):

1. In sparse binary tables, the empirical distribution of the overall tests G^2 and X^2 does not match its asymptotic distribution. Therefore, statistical inferences based on these statistics are invalid in sparse tables. Although it is possible to generate the empirical sampling distribution of these statistics using resampling methods (for instance, using parametric bootstrap; Bartholomew and Tzamourani, 1999), the amount of computation involved is substantial, particularly when we are interested in comparing the fit of competing IRT models to data sets with a large number of variables.
2. When G^2 and X^2 indicate a poorly fitting model, one is interested in identifying the source of the misfit. Because the number of cell residuals to be inspected is generally very large, it is difficult if not impossible to draw useful information about the source of the misfit using cell residuals (Bartholomew & Knott, 1999). In recent years it has been advocated (e.g., Bartholomew & Tzamourani, 1999; McDonald & Mok, 1995; Reiser, 1996) to inspect low-order marginal residuals (e.g., univariate, bivariate, and trivariate residuals) to detect the source of any possible misfit. Although it is not difficult to derive the asymptotic distribution of low-order marginal residuals, no overall limited information tests with known asymptotic distribution seemed to be available in the item response modeling literature (but see Bartholomew & Leung, 2001; Reiser, 1996).
3. Several limited-information estimation procedures have been proposed to estimate item response models (e.g., Christoffersson, 1975; McDonald, 1982b; Muthén, 1978; see also Maydeu-Olivares, 2001). These procedures yield limited-information goodness-of-fit tests of known asymptotic distribution that perform well in sparse tables (Maydeu-Olivares, 2001). However, when limited information estimation procedures are used, G^2 and X^2 do not follow their

usual asymptotic distribution (Bishop, Feinberg, & Holland, 1975), and no full-information goodness-of-fit test with known asymptotic distribution had been proposed for these estimators.

Maydeu-Olivares and Joe (in press) recently addressed these challenges by introducing a unified framework for limited- and full-information estimation and testing in binary contingency tables using the joint raw moments of the MVB distribution. These moments can be expressed as a linear function of the cell probabilities $\dot{\boldsymbol{\pi}} = \mathbf{T}\boldsymbol{\pi}$, where $\mathbf{T}$ is a matrix that consists of ones and zeros. Consider now partitioning the vector of joint raw moments of the MVB distribution as $\dot{\boldsymbol{\pi}}' = (\dot{\boldsymbol{\pi}}_1, \dot{\boldsymbol{\pi}}_2, \ldots, \dot{\boldsymbol{\pi}}_n)'$, where $\dot{\boldsymbol{\pi}}_i$ denotes the $\binom{n}{i}$-dimensional vector of ith-order moments (see the Appendix). $\mathbf{T}$ can also be partitioned according to the partitioning of $\dot{\pi}$ as $\mathbf{T} = (\mathbf{T}_1', \mathbf{T}_2', \ldots, \mathbf{T}_n')'$, where $\mathbf{T}_i$ is a $\binom{n}{i} \times 2^n$ matrix of ones and zeros (see the example shown in Equation A3). Consider now the $s = \sum_{i=1}^{r} \binom{n}{i}$, dimensional vector of moments up to order $r \le n$ $\tilde{\boldsymbol{\pi}}_r' = (\dot{\boldsymbol{\pi}}_1, \dot{\boldsymbol{\pi}}_2, \ldots, \dot{\boldsymbol{\pi}}_r)'$, with sample counterpart $\tilde{\mathbf{p}}_r$. Letting $\tilde{\mathbf{T}}_r = (\mathbf{T}_1', \mathbf{T}_2', \ldots, \mathbf{T}_r')'$, we can write

$$\tilde{\boldsymbol{\pi}}_r = \tilde{\mathbf{T}}_r \boldsymbol{\pi}. \tag{45}$$

Then, from Equations 45 and 37, it follows immediately that the asymptotic distribution of the joint sample raw moments up to order r of the MVB distribution is

$$\sqrt{N}\,(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r) \xrightarrow{d} N(\mathbf{0}, \tilde{\boldsymbol{\Xi}}_r), \qquad \tilde{\boldsymbol{\Xi}}_r = \tilde{\mathbf{T}}_r \boldsymbol{\Gamma} \tilde{\mathbf{T}}_r. \tag{46}$$

Using this result, Maydeu-Olivares and Joe (in press) proposed a unifying framework for limited- and full-information testing in binary contingency tables using

$$M_r = N(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\hat{\boldsymbol{\theta}}))' \hat{\mathbf{U}}_r (\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\hat{\boldsymbol{\theta}})), \tag{47}$$

$$\boldsymbol{U}_r = \tilde{\boldsymbol{\Xi}}_r^{-1} - \tilde{\boldsymbol{\Xi}}_r^{-1} \tilde{\boldsymbol{\Delta}}_r (\tilde{\boldsymbol{\Delta}}_r' \tilde{\boldsymbol{\Xi}}_r^{-1} \tilde{\boldsymbol{\Delta}}_r)^{-1} \tilde{\boldsymbol{\Delta}}_r' \tilde{\boldsymbol{\Xi}}_r^{-1}, \tag{48}$$

where $\tilde{\boldsymbol{\Delta}}_r = \partial \tilde{\boldsymbol{\pi}}_r(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$, and all matrices are evaluated at the estimated parameter values. Maydeu-Olivares and Joe showed that if $\boldsymbol{\theta}$ is estimated using any (limited or full information) consistent and asymptotically normal estimator and if $\tilde{\boldsymbol{\Delta}}_r$ is of full rank q (i.e., if the model is locally identified from the moments up to order r), then M_r is asymptotically distributed as a chi-square with $s - q$ degrees of freedom.

M_r the moments of the binaries data up to order r to assess the goodness of fit of the model. Its limiting case, M_n is a full-information statistic because of the one-to-one relation between the set of all marginal moments and the cell probabilities in Equations A4 and A5. Furthermore, Maydeu-Olivares and Joe showed that M_n can be alternatively be written as

$$M_n = N(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}))'\hat{\mathbf{U}}(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})), \qquad \mathbf{U} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{D}^{-1}, \tag{49}$$

where all the matrices are to be evaluated at the estimated parameter values. This statistic is asymptotically distributed as a chi-square with $2^n - q - 1$ degrees of freedom for any consistent and asymptotically normal estimator (including limited-information estimators). Also, $M_n = X^2$ when the model is estimated by full-information maximum likelihood.

It is interesting to point out that when applied to binary data, the statistic T_B proposed by Browne (1984) in the context of covariance structure modeling is closely related to the member of the class of Equation 47 where only univariate and bivariate moments are used (i.e., M_2). In fact, M_2 is asymptotically equal to the T_B statistic. Both statistics are asymptotically chi-square distributed with $t - q$ degrees of freedom for any consistent and asymptotically normal estimator. However, they are not algebraically equal. M_2 is a quadratic form in residual raw univariate and bivariate moments, whereas T_B is a quadratic form in residual covariances (bivariate central moments). Furthermore, the asymptotic covariance matrix of the sample moments used in each statistic is estimated differently. In M_2 this matrix is evaluated at the estimated parameter values, whereas in T_B it is estimated using sample moments. Nevertheless, it is remarkable that since Browne's T_B statistic was proposed in 1984, no one seems to have noticed that if an IRT model is identified from the univariate and bivariate margins, then the T_B statistic can be used to test the goodness of fit of the model.

In closing this discussion on goodness-of-fit statistics, we present an alternative family of test statistics, M_r', introduced by Maydeu-Olivares and Joe (in press), which can also be used to assess the goodness of fit of IRT models and has a greater resemblance to Browne's statistic. This family is

$$M_r' = N(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\hat{\boldsymbol{\theta}}))'\hat{\mathbf{U}}_r'(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\hat{\boldsymbol{\theta}})), \tag{50}$$

where $\hat{\mathbf{U}}_r'$ denotes Equation 48 evaluated as in Browne's statistic, that is, the derivative matrices are evaluated at the estimated parameter values, but $\tilde{\boldsymbol{\Xi}}_r$ is evaluated using sample proportions. Obviously $M_r' \overset{a}{=} M_r \overset{d}{\to} \chi^2_{s-q}$.

In a similar fashion, a unifying framework for limited and full information estimation of IRT models for binary data can be obtained using quadratic forms in joint raw moments of the MVB distribution. Consider the fit function (Maydeu-Olivares & Joe, in press)

$$F_r = (\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\boldsymbol{\theta}))'\,\hat{\mathbf{W}}_r\,(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\boldsymbol{\theta}))\,, \tag{51}$$

where $\hat{\mathbf{W}}_r$ is a matrix converging in probability to $\mathbf{W}_r$, a positive-definite matrix that does not depend on $\boldsymbol{\theta}$. Some obvious choices for $\hat{\mathbf{W}}_r$ in Equation 51 are $\hat{\mathbf{W}}_r = \mathbf{I}$,

$\hat{\mathbf{W}}_r = (\text{Diag}(\hat{\tilde{\boldsymbol{\Xi}}}_r))^{-1}$, and $\hat{\mathbf{W}}_r = \hat{\tilde{\boldsymbol{\Xi}}}_r^{-1}$, where $\hat{\tilde{\boldsymbol{\Xi}}}_r$ denotes $\tilde{\boldsymbol{\Xi}}_r$ consistently estimated using sample proportions. If $\tilde{\boldsymbol{\Delta}}_r$ is of full rank q and some other mild regularity conditions are satisfied, $\hat{\boldsymbol{\theta}}$ obtained by minimizing Equation 51 is consistent and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{H}\tilde{\boldsymbol{\Xi}}_r\mathbf{H}'), \qquad \mathbf{H} = (\tilde{\boldsymbol{\Delta}}_r'\mathbf{W}_r\tilde{\boldsymbol{\Delta}}_r)^{-1}\tilde{\boldsymbol{\Delta}}_r'\mathbf{W}_r, \tag{52}$$

$$\sqrt{N}(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\hat{\boldsymbol{\theta}})) \xrightarrow{d} N(\mathbf{0}, \tilde{\boldsymbol{\Omega}}_r), \qquad \tilde{\boldsymbol{\Omega}}_r = (\mathbf{I} - \tilde{\boldsymbol{\Delta}}_r\mathbf{H})\tilde{\boldsymbol{\Xi}}_r(\mathbf{I} - \tilde{\boldsymbol{\Delta}}_r\mathbf{H})'. \tag{53}$$

To obtain standard errors for the parameter estimates and residual proportions, the derivative matrices may be evaluated at the estimated parameter values, and $\tilde{\boldsymbol{\Xi}}_r$ may be evaluated using sample proportions. Note that when F_n is employed, a class of full-information estimators is obtained. Maydeu-Olivares and Joe (in press) explicitly related the class of estimators F_n to the class of minimum chi-square estimators in Equation 42.

When $\hat{\mathbf{W}}_r = \hat{\tilde{\boldsymbol{\Xi}}}_r^{-1}$ is used in Equation 51, Equations 52 and 53 simplify to

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0}, \left(\tilde{\boldsymbol{\Delta}}_r'\tilde{\boldsymbol{\Xi}}_r^{-1}\tilde{\boldsymbol{\Delta}}_r\right)^{-1}\right) \tag{54}$$

and

$$\sqrt{N}(\tilde{\mathbf{p}}_r - \tilde{\boldsymbol{\pi}}_r(\hat{\boldsymbol{\theta}})) \xrightarrow{d} N\left(\mathbf{0}, \tilde{\boldsymbol{\Xi}}_r - \tilde{\boldsymbol{\Delta}}_r\left(\tilde{\boldsymbol{\Delta}}_r'\tilde{\boldsymbol{\Xi}}_r^{-1}\tilde{\boldsymbol{\Delta}}_r\right)^{-1}\tilde{\boldsymbol{\Delta}}_r'\right). \tag{55}$$

respectively, and we obtain estimators that are asymptotically efficient among the class of estimator using information up to order r. Furthermore,

$$N\hat{F}_r = M_r' \xrightarrow{d} \chi^2_{s-q}. \tag{56}$$

The estimator proposed by Christoffersson (1975) to estimate the normal ogive model is a member of the family of estimators (51). He estimated the model minimizing $F_2 = (\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta}))'\hat{\tilde{\boldsymbol{\Xi}}}_2^{-1}(\tilde{\mathbf{p}}_2 - \tilde{\boldsymbol{\pi}}_2(\boldsymbol{\theta}))$.

NUMERICAL EXAMPLES

We provide two numerical examples to illustrate our discussion using the Law School Admissions Test (LSAT) 6 and LSAT 7 data sets (Bock & Lieberman, 1970). Each of these data sets consists of 1,000 observations on five binary variables.

Comparing the Fit of a Factor Model and of a Logistic Model to the LSAT 6 Data Using Browne's T_B Statistic

In this section we compare the fit of a factor model versus a logistic IRT model applied to the LSAT 6 data. We discussed previously that Browne's T_B statistic can

P1: FCH/FFX P2: FCH/FFX QC: FCH/FFX T1: FCH
LE104-03 LE104-Maydeu-v2.cls February 20, 2005 13:49

be used to this purpose. We estimated a one-factor model to the LSAT 6 data using unweighted least squares using the sample covariances under the α parametrization in Equation 26. The two-parameter logistic IRT model was estimated using maximum likelihood from Equation 35. The parameter estimates, standard errors, and T_B statistics are shown in Table 3.1. We do not observe that the logistic IRT model outperforms the factor model in fitting these data, as assessed by the T_B statistic.

Notice that Table 3.1 does not report any of the IRT goodness-of-fit statistics for the factor model. This is because under the factor model assumptions in Equations 18 and 19 these statistics cannot be computed. Additional assumptions on third- and higher order moments of the latent trait are needed to compute the expected probabilities under the factor model. These expected probabilities are needed to compute the IRT fit statistics.

Factor Modeling Versus Linear IRT Modeling of the LSAT 6 Data

In the factor model only univariate and bivariate moments are specified. Therefore, this model can only be estimated using univariate and bivariate information. Unlike the factor model, in the linear IRT model all the moments of the latent traits are specified. As a result, the linear IRT model can be estimated using either full information or limited information. Here, we compare the fit of a unidimensional linear IRT model versus the fit of a one-factor model applied to the LSAT 6 data. We assume that the moments of the latent trait in the linear IRT model are fixed constants. The constants chosen are those of a standard normal density. Therefore, the five moments of the latent trait are fixed to

$$\boldsymbol{\kappa}' = (0, 1, 0, 3, 0)\,. \tag{57}$$

Table 3.1 reports the linear IRT parameters estimated using a variety of full- and limited-information estimators.

Because the factor model and the linear IRT model are equivalent when the latter is estimated using only bivariate information, it is most interesting to compare the last two columns of Table 3.1, where both models are estimated using bivariate information. The results are not identical even though we used the same estimation procedure (ULS). This is because the linear IRT model is estimated from raw moments (marginal proportions), whereas the factor model is estimated using central moments (covariances) and there is not a one-to-one correspondence between both fit functions.

Effects of the Estimation Method and Choice of IRT Model on the LSAT 6 Data

In Table 3.1 we present the results of fitting a two-parameter logistic IRT model to the LSAT 6 data using (a) full-information maximum likelihood and

TABLE 3.1
Parameter Estimates and Goodness-of-Fit Tests for the Law School Admission Test 6 Data

	Logistic IRT Model		*Linear IRT Model*			*Factor Model*
	ML (Full)	*ULS (Bivariate)*	*ML (Full)*	*ULS (Full)*	*ULS (Bivariate)*	*ULS (Bivariate)*
Parameter estimates						
α_1	2.77	2.84	0.92	0.92	0.92	0.92
	(0.21)	(0.24)	(0.01)	(0.01)	(0.01)	(0.01)
α_2	0.99	0.98	0.71	0.71	0.71	0.71
	(0.09)	(0.09)	(0.01)	(0.02)	(0.01)	(0.01)
α_3	0.25	0.26	0.55	0.56	0.55	0.55
	(0.08)	(0.08)	(0.02)	(0.02)	(0.02)	(0.03)
α_4	1.28	1.27	0.76	0.77	0.76	0.76
	(0.10)	(0.10)	(0.01)	(0.01)	(0.01)	(0.01)
α_5	2.05	2.03	0.87	0.87	0.87	0.87
	(0.13)	(0.12)	(0.01)	(0.01)	(0.01)	(0.01)
β_1	0.83	0.97	0.05	0.05	0.07	0.06
	(0.26)	(0.31)	(0.02)	(0.02)	(0.02)	(0.02)
β_2	0.72	0.63	0.14	0.14	0.12	0.13
	(0.18)	(0.20)	(0.03)	(0.04)	(0.03)	(0.04)
β_3	0.89	1.04	0.18	0.17	0.21	0.19
	(0.23)	(0.36)	(0.04)	(0.05)	(0.05)	(0.03)
β_4	0.69	0.62	0.12	0.10	0.11	0.12
	(0.18)	(0.20)	(0.03)	(0.03)	(0.03)	(0.03)
β_5	0.66	0.55	0.07	0.06	0.06	0.07
	(0.20)	(0.22)	(0.02)	(0.02)	(0.02)	(0.02)
Goodness-of-fit tests						
X^2	18.15	19.62	19.51	20.69	22.47	—
	(0.64)	—	(0.55)	—	—	
G^2	21.47	22.49	22.96	23.96	25.66	—
	(0.43)	—	(0.35)	—	—	
M_n	18.15	18.79	19.51	19.68	19.69	—
	(0.64)	(0.60)	(0.55)	(0.54)	(0.54)	
M_2	4.75	5.07	4.37	4.49	4.70	—
	(0.45)	(0.41)	(0.50)	(0.48)	(0.45)	
T_B	5.06	5.37	4.89	4.83	5.20	4.90
	(0.41)	(0.37)	(0.43)	(0.42)	(0.39)	(0.43)

Note. IRT, Item response theory. Estimators are maximum likelihood (ML) or unweighted least squares (ULS). Information is full or bivariate as indicated. The factor model and the linear item response model estimated from bivariate information are equivalent models. Standard errors are given in parentheses for parameter estimates; p values are given in parentheses for goodness-of-fit tests. When the model is not estimated by full-information maximum likelihood, p values for X^2 and G^2 are not provided because these statistics are not asymptotically chi-squared distributed. There are 21 degrees of freedom for X^2, G^2, and M_n; there are 5 degrees of freedom for M_2 and T_B.

(b) bivariate-information ULS estimation. We also present the results of fitting a linear IRT model to these data using (a) full-information maximum likelihood, (b) full-information ULS estimation, and (c) bivariate-information ULS estimation. Therefore, we can examine the effects of the choice of model and of the choice of estimation method. We can use three statistics to compare the fit of these two models regardless of how they have been estimated: the full-information statistic M_n and the limited-information statistics M_2 and T_B.

Informally speaking, the M_n statistic can be used to assess the fit of these models to the cell proportions. The M_2 statistic can be used to assess their fit to the univariate and bivariate raw moments of the data. Finally, the T_B statistic can be used to assess their fit to the sample covariances. When the model is estimated using full-information maximum likelihood, $M_n = \mathrm{X}^2$. Also, when the model is not estimated using an asymptotically efficient estimator, X^2 and G^2 are not asymptotically chi-square distributed and consequently p values are not reported in Table 3.1 in those instances.

Inspecting the relevant goodness-of-fit statistics presented in this table, we see that for these data the difference between estimating a model using full-information maximum likelihood versus bivariate-information ULS is very small. Also, the fit differences between the linear and the logistic models, for these data are also rather small. In general, one should expect the logistic model to yield a better fit to binary data than the linear model (see the next example), but for these data the logistic item response functions are so flat that the linear item response model provides a comparable fit. This is illustrated in Fig. 3.1, where we provide the item response functions under both models for a chosen item.

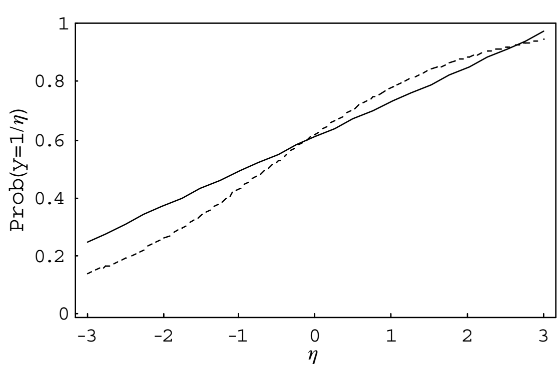

FIG. 3.1. Item response function of the Law School Admission Test 6 fourth item under the linear and logistic models. The parameter estimates were estimated using full-information maximum likelihood. The parameter estimates are depicted in Table 3.1.

Linear Versus Logistic IRT Modeling of the LSAT 7 Data

We now examine the fit of unidimensional linear and logistic IRT models to the LSAT 7 data. Here we only used full-information maximum likelihood estimation. Initially, we fixed the moments of the latent trait in the linear IRT model at the values of the moments of a standard normal density (see Equation 57). The results are shown in Table 3.2. As can be seen in this table, there is not much difference

TABLE 3.2
Parameter Estimates and Goodness-of-Fit Tests for the Law School Admission Test 7 Data

	Logistic Model	*Linear Model A*	*Linear Model B*
Parameter estimates			
α_1	1.86	0.83	0.79
	(0.13)	(0.01)	(0.02)
α_2	0.81	0.66	0.60
	(0.09)	(0.01)	(0.02)
α_3	1.81	0.77	0.71
	(0.20)	(0.01)	(0.02)
α_4	0.49	0.60	0.56
	(0.07)	(0.02)	(0.02)
α_5	1.85	0.84	0.82
	(0.11)	(0.01)	(0.02)
β_1	0.99	0.12	0.14
	(0.17)	(0.02)	(0.02)
β_2	1.08	0.20	0.20
	(0.17)	(0.02)	(0.02)
β_3	1.71	0.20	0.23
	(0.32)	(0.02)	(0.03)
β_4	0.77	0.14	0.17
	(0.13)	(0.02)	(0.02)
β_5	0.74	0.08	0.10
	(0.15)	(0.02)	(0.02)
κ_1	0	0	0.28
	(Fixed)	*(Fixed)*	(0.07)
Goodness-of-fit tests			
X^2	32.48	46.56	34.09
	(0.05)	(<0.01)	(0.03)
G^2	31.94	42.98	32.11
	(0.06)	(<0.01)	(0.04)
M_2	11.92	10.19	11.27
	(0.04)	(0.07)	—

Note. Standard errors are given in parentheses for parameter estimates; p values are given in parentheses for goodness of fit tests. All models were estimated by full-information maximum likelihood. The number of degrees of freedom for X^2 and G^2 is 21 for the logistic model and linear model A and 20 for linear model B. The number of degrees of freedom for M_2 is 5 for the logistic model and linear model A. The values used to fix the latent variable moments were those of a standard normal density.

TABLE 3.3
Goodness-of-Fit Tests for Some Unidimensional Linear Item Response Models Applied to the Law School Admission Test 7 Data

Model	G^2	*df*
All moments fixed	42.98	21
Estimated κ_1	32.11	20
Estimated κ_2	41.86	20
Estimated $\{\kappa_1, \kappa_2\}$	31.76	19
Estimated $\{\kappa_1, \kappa_2, \kappa_3\}$	31.50	18

Note. All models were estimated by full-information maximum likelihood. The values used to fix the moments were those of a standard normal density.

in how well both models reproduce the bivariate margins of the table. However, the full-information test statistics indicate that the linear model fails to reproduce the observed cell frequencies. An examination of the standardized cell residuals $N(p_c - \pi_c)^2/\pi_c$, where π_c denotes a cell probability as in Equation 1, reveals that the linear model particularly fails to reproduce the patterns (0,1,0,0,0) and (0,0,0,0,0). Their corresponding standardized cell residuals are 13.81 and 7.81, respectively. Thus, these patterns account for 28% and 17%, respectively, of the value of the X^2 statistic.

However, we can improve the fit of the linear IRT model by estimating some of the moments of the latent trait. With five items, up to three moments can be identified. In Table 3.3 we provide the values of the G^2 statistics obtained when some of the moments of the latent trait were estimated. As can be seen in this table, the best unidimensional linear model for these data is obtained by estimating the mean of the latent trait. In Table 3.2 we provide the full set of parameter estimates and standard errors for this model. This model provides a fit to the LSAT 7 data comparable to that of the logistic model, at the expense of an additional parameter. Note that we do not provide a p value for M_2 because this model is not identified from bivariate information.

It should be noted that estimating a high-order moment of a random variable requires large samples, more so, probably, in the case of latent variables. Thus, estimating high-order moments of a latent trait should only be attempted in large samples. If the sample size is not large enough, the linear model may become empirically underidentified (i.e., $\hat{\boldsymbol{\Delta}}$ will not be of full rank).

EAP Scores for the Linear Model

Once the parameters of a linear model have been estimated, we can obtain scores for individual responses. Here we compare the results obtained when computing expected a posteriori scores for the estimated-mean linear model and for the logistic model for the LSAT 7 dataset. The parameter estimates for these models were presented in Table 3.2. For the logistic model, EAP scores were computed Equation

12 assuming a prior standard normal density because this is the density we used in estimating the parameters of this model.

For the linear model, although it is not necessary to assume a density for the latent traits to estimate the model parameters, it is necessary to use some prior distribution to obtain the posterior distribution of the latent traits. In the unidimensional case, we have found that the normal prior distribution

$$\gamma_1(\eta) = \phi_1\left(\eta\colon -\kappa_1, \kappa_2 - \kappa_1^2\right) \tag{58}$$

yields good results. When the EAP scores for the mean-estimated linear model are obtained using this prior distribution, they correlate 0.98 with the logistic EAP scores and 0.95 with the number right scores (i.e., the unweighted sum of the binary scores). Figures 3.2 and 3.3 are plots of the linear EAP scores against the logistic EAP and number-right scores.

Similar results were obtained when we computed EAP scores for the LSAT 6 data using the linear and logistic models estimated by full-information maximum likelihood. The linear EAP scores correlated 0.96 with the logistic EAP scores, and 0.96 with the number-right scores.

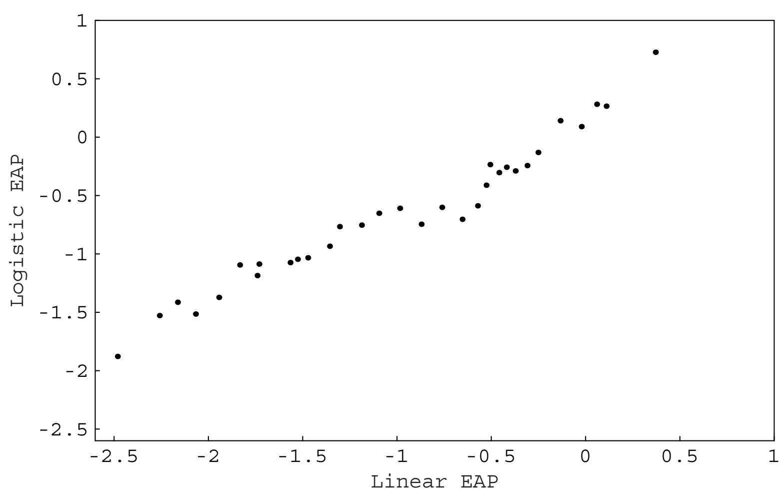

FIG. 3.2. Plot of the expected a posterior (EAP) latent trait estimates under the logistic model and a linear model estimating the mean of the latent trait for the Law School Admission Test 7 data. The parameter estimates were estimated using full-information maximum likelihood. The parameter estimates are depicted in Table 3.2.

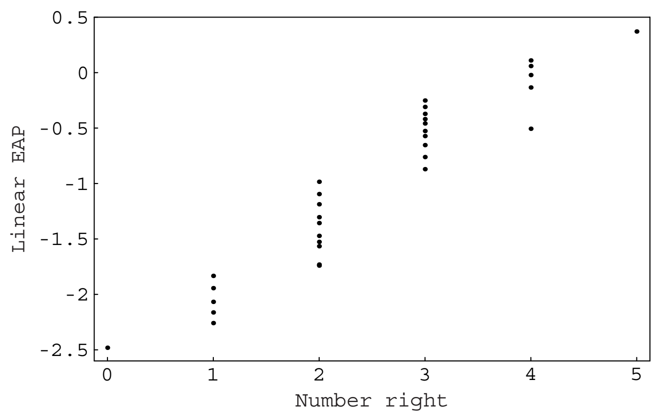

FIG. 3.3. Plot of the expected a posterior (EAP) latent trait estimates under a linear model estimating the mean of the latent trait versus number-right score for the Law School Admission Test 7 data.

DISCUSSION AND CONCLUSIONS

In this chapter we integrated factor analysis and IRT within a unified modeling and estimation framework. McDonald (1982a) offered a unified treatment of linear and nonlinear IRT modeling using Hermite polynomials. He also offered an alternative unified treatment of linear and nonlinear IRT modeling using link functions (McDonald, 1999). When the latter approach is employed, it is not clear what the difference is, if any, between a factor model applied to binary data and a linear item response model for binary data. We discussed that these two models differ in that the linear item response model is a fully specified model, whereas under the factor model only the first two moments of the data are specified. As a result, under the linear item response model, most moments of the latent traits can be identified when full-information estimation is used.

One attractive feature of the linear item response model is that it does not require any assumptions on the distribution of the latent traits. Only assumptions on the moments of the latent traits are needed to identify the model. Furthermore, we showed that high-order moments of the latent traits can be estimated under the linear IRT model. We illustrated this fact by estimating the first three moments of the latent trait along with the item parameters of a unidimensional model fitted to the LSAT 7 data. Note, however, that large samples are needed to estimate high-order moments of any random variable, more so, probably in the case of latent random variables. Also note that although we have not assumed any prior distribution of the latent traits to estimate the item parameters, we needed to assume a prior distribution of

P1: FCH/FFX P2: FCH/FFX QC: FCH/FFX T1: FCH
LE104-03 LE104-Maydeu-v2.cls February 20, 2005 13:49

the latent traits to obtain the posterior distribution of the latent traits, which is the approach taken here to compute individual scores under the linear IRT model.

An unattractive feature of the linear model is that the item response functions are not bounded between zero and one. Thus, in principle, for low enough values of the latent traits the probability of endorsing the item may be negative, whereas for high enough values of the latent traits the probability of endorsing the item may be greater than one. However, we verified that at the EAP scores computed for the LSAT 6 and LSAT 7 data the item response functions do not fall outside of the [0, 1] range. Also, for these two data sets the linear item response model is a proper model because the cell probabilities are in the range [0, 1].

Although in general we expect nonlinear IRT models to yield a better fit to binary data than the linear model, we also showed using two data sets that in some applications the linear model may provide a good fit to binary data sets. For the LSAT 6 data, a linear model with fixed moments provides a fit comparable to that of the two-parameter logistic model. For the LSAT 7 data, a linear model with fixed moments provides a poor fit to the observed binary pattern frequencies, but a linear model estimating the mean of the latent traits provides a fit comparable to that of a two-parameter logistic model (at the expense of an additional parameter, of course).

In closing, we note that McDonald (1999) pointed out that when the linear IRT model is estimated using only univariate and bivariate information, this model is equivalent to the factor model applied to binary data. However, the factor model is generally estimated using central joint moments (covariances) or standardized joint central moments (correlations), whereas in limited-information IRT estimation raw joint moments (cross-products) are generally used. In any case, the general framework of moment estimators provides a unifying estimation framework for factor analysis (and more generally structural equation modeling) and IRT. We pointed out that Browne's T_B statistic provides a common yardstick for assessing the goodness of fit of a factor model and an IRT model to binary data. This statistic is a quadratic form in the residual covariances with a sample-based weight matrix. Maydeu-Olivares and Joe (in press) recently introduced a similar statistic, M_2. This is also a quadratic form, but in the residual cross-products, where a model-based weight matrix is used instead. The two statistics are asymptotically chi-square distributed for any consistent and asymptotically normal estimator, and so is the full information extension of M_2, M_n. Because M_n is also asymptotically chi-square distributed for any consistent and asymptotically normal estimator it can be used, unlike X^2 or G^2, to assess the goodness of fit of competing IRT models regardless of whether they have been estimated using limited- or full-information methods.

APPENDIX. REPRESENTATIONS OF THE MULTIVARIATE BERNOULLI DISTRIBUTION

The $\binom{n}{1}$ univariate moments of the multivariate Bernoulli distribution are of the form

$$\mathrm{E}[y_i] = 1 \times \Pr(y_i = 1) + 0 \times \Pr(y_i = 0) = \Pr(y_i = 1) = \dot{\pi}_i. \tag{A1}$$

Similarly, each of the $\binom{n}{2}$ bivariate raw moments of $\mathbf{y}$ is of the form

$$\mathrm{E}[y_i y_j] = \Pr[(y_i = 1) \cap (y_j = 1)] = \dot{\pi}_{ij}, \quad i < j. \tag{A2}$$

and so forth. The overall number of raw joint moments of $\mathbf{y}$ is $\sum_{i=1}^{n} \binom{n}{i} = 2^n - 1$. The relationship between the $(2^n - 1)$ vector of moments $\dot{\boldsymbol{\pi}}$ and the 2^n vector of cell probabilities $\boldsymbol{\pi}$ is linear, say $\dot{\boldsymbol{\pi}} = \mathbf{T}\boldsymbol{\pi}$, where $\mathbf{T}$ is a matrix that consists of ones and zeros (Maydeu-Olivares, 1997).

We illustrate $\dot{\boldsymbol{\pi}} = \mathbf{T}\boldsymbol{\pi}$ for the case of $n = 3$ Bernoulli variables:

$$\begin{pmatrix} \dot{\pi}_1 \\ \dot{\pi}_2 \\ \dot{\pi}_3 \\ \dot{\pi}_{12} \\ \dot{\pi}_{13} \\ \dot{\pi}_{23} \\ \dot{\pi}_{123} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_{000} \\ \pi_{100} \\ \pi_{010} \\ \pi_{001} \\ \pi_{110} \\ \pi_{101} \\ \pi_{011} \\ \pi_{111} \end{pmatrix}, \tag{A3}$$

where, for instance, $\pi_{100} = \Pr[(y_1 = 1) \cap (y_2 = 0) \cap (y_3 = 0)]$.

The relationship between $\boldsymbol{\pi}$ and $\dot{\boldsymbol{\pi}}$ is one-to-one. To see this, notice in Equation A3 that $\dot{\boldsymbol{\pi}} = \mathbf{T}\boldsymbol{\pi}$ can always be written as

$$\dot{\boldsymbol{\pi}} = (\mathbf{0} \ \breve{\mathbf{T}}) \begin{pmatrix} \pi_0 \\ \breve{\boldsymbol{\pi}} \end{pmatrix} = \breve{\mathbf{T}}\breve{\boldsymbol{\pi}}, \tag{A4}$$

where $\pi_0 = \Pr[\bigcap_{i=1}^{n} (y_i = 0)]$, $\breve{\boldsymbol{\pi}}$ is used to denote the $(2^n - 1)$-dimensional vector of cell probabilities excluding π_0, and $\breve{\mathbf{T}}$ is an upper triangular square matrix. Then, because $\pi_0 = 1 - \mathbf{1}'\breve{\boldsymbol{\pi}}$, the inverse relationship between $\dot{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}$ is

$$\pi = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} -\mathbf{1}'\breve{\mathbf{T}}^{-1} \\ \breve{\mathbf{T}}^{-1} \end{pmatrix} \dot{\boldsymbol{\pi}}. \tag{A5}$$

As a result, we can represent any item response model for binary data using its vector of moments $\dot{\boldsymbol{\pi}}$ rather than its vector of cell probabilities $\boldsymbol{\pi}$.

We now present some results for the moments of the multivariate Bernoulli distribution that are valid under any item response model. We make use of these results in the body of the text.

First, we notice that the expected value of a variable given the latent traits simply equals the item response function. This is because

$$\mathrm{E}\left[y_i|\boldsymbol{\eta}\right] = 1 \times \Pr\left(y_i = 1|\boldsymbol{\eta}\right) + 0 \times \Pr\left(y_i = 0|\boldsymbol{\eta}\right) = \Pr\left(y_i = 1|\boldsymbol{\eta}\right). \quad \text{(A6)}$$

Next, we notice that the univariate moments are simply the expected value of the item response function,

$$\dot{\pi}_i = \mathrm{E}_{\boldsymbol{\eta}}\left[\Pr\left(y_i = 1|\boldsymbol{\eta}\right)\right], \quad \text{(A7)}$$

where $\mathrm{E}_{\boldsymbol{\eta}}[\bullet]$ is used to indicate that the expectation is to be taken with respect to $\boldsymbol{\eta}$. This result follows immediately from Equation A6 and the double expectation theorem (e.g., Mittelhammer, 1996), $\mathrm{E}[y_i] = \mathrm{E}_{\boldsymbol{\eta}}[\mathrm{E}[y_i|\boldsymbol{\eta}]]$.

Similarly, we notice that the bivariate raw moments are simply

$$\dot{\pi}_{ij} = \mathrm{E}_{\boldsymbol{\eta}}[\Pr(y_i = 1|\eta)\Pr(y_j = 1|\eta)]. \quad \text{(A8)}$$

This is because we can write $\dot{\pi}_{ij} = \mathrm{E}[y_i \cap y_j] = \mathrm{E}_{\boldsymbol{\eta}}[\mathrm{E}[(y_i \cap y_j)|\boldsymbol{\eta}]]$. From the assumption of local independence, however, $\mathrm{E}[(y_i \cap y_j)|\boldsymbol{\eta}] = \mathrm{E}[y_i|\boldsymbol{\eta}]\mathrm{E}[y_j|\boldsymbol{\eta}]$. Finally, the trivariate moments are simply

$$\dot{\pi}_{ijk} = \mathrm{E}_{\eta}[\Pr(y_i = 1|\boldsymbol{\eta})\Pr(y_j = 1|\boldsymbol{\eta})\Pr(y_k = 1|\boldsymbol{\eta})]. \quad \text{(A9)}$$

Similar expressions result for higher moments.

ACKNOWLDGMENTS

This research was supported by grant BSO 2000-0661 of the Spanish Ministry of Science and Technology and by a Distinguished Young Investigator Award of the Catalan government.

REFERENCES

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

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

Bartholomew, D. J., & Leung, S. O. (2001). A goodness of fit test for sparse $2p$ contingency tables. *British Journal of Mathematical and Statistical Psychology*, *55*, 1–16.

Bartholomew, D. J., & Tzamourani, P. (1999). The goodness of fit of latent trait models in attitude measurement. *Sociolological Methods and Research*, *27*, 525–546.

Bekker, P. A., Merckens, A., & Wansbeek, T. J. (1994). *Identification, equivalent models and computer algebra*. San Diego, CA: Academic.

Bishop, Y., Feinberg, S. E., & Holland, P. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika*, *35*, 179–197.

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

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

Ferguson, T. S. (1996). *A course in large sample theory*. London: Chapman & Hall.

Maydeu-Olivares, A. (1996). Modelos multidimensionales de respuesta a los items [Multidimensional item response models]. In J. Muñiz (Ed.), *Psicometría* [Psychometrics] (pp. 811–868). Madrid: Universitas.

Maydeu-Olivares, A. (1997). Structural equation modeling of binary preference data. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1997). *Dissertation Abstracts International*, *58*, 5694B.

Maydeu-Olivares, A. (2001). 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., & Joe, H. (in press). Limited information estimation and testing in sparse binary contingency tables: A unified framework. *Journal of the American Statistical Association.*

Maydeu-Olivares, A., Hernández, A., & McDonald, R. P. (2002). *A multidimensional unfolding latent trait model for binary attitudinal data*. Manuscript under review. Barcelona: University of Barcelona.

McDonald, R. P. (1967). *Nonlinear factor analysis*. (Psychometric Monographs, No. 15). Chicago: University of Chicago Press.

McDonald, R. P. (1969). The common factor analysis of multicategory data. *British Journal of Mathematical and Statistical Psychology*, *22*, 165–175.

McDonald, R. P. (1980). A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *British Journal of Mathematical and Statistical Psychology*, *33*, 161–183.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100–117.

McDonald, R. P. (1982a). Linear vs. nonlinear models in latent trait theory. *Applied Psychological Measurement*, *6*, 379–396.

McDonald, R. P. (1982b). Fitting latent trait models. In D. Spearrit (Ed.), *Some alternative approaches to the improvement of measurement in education and psychology* (pp. 213–237). Hawthorn: Australian Council for Educational Research.

McDonald, R. P. (1985a). Unidimensional and multidimensional models for item response theory. In D. J. Weiss (Ed.), *Proceedings of the 1982 Item Response and Computerized Adaptive Testing Conference* (pp. 127–147). Minneapolis: University of Minnesota Press.

McDonald, R. P. (1985b). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McDonald, R. P. (1986). Describing the elephant: Structure and function in multivariate data. *Psychometrika*, *51*, 513–534.

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 approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McDonald, R. P. (2001). A basis for multidimensional item response theory. *Applied Psychological Measurement*, *24*, 99–114.

McDonald, R. P., & Mok, M. C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research*, *54*, 483–495.

Mittelhammer, R. C. (1996). *Mathematical statistics for economics and business*. New York: Springer.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, *65*, 391–411.

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

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika*, *61*, 509–528.

Reiser, M., & VandenBerg, M. (1994). Validity of the chi-square test in dichotomous varible factor analysis when expected frequencies are small. *British Journal of Mathematical and Statistical Psychology*, *47*, 85–107.

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

van der Linden, W. J., & Hambleton, R. K. (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.

In the framework of response theory the lowest order contribution to the parity violating potential VPV is given by the linear response function (denoted by. ã€ˆã€ˆ.;.ã€‰ã€‰Ï‰1). ã€ˆã€ˆ.;.ã€‰ã€‰Ï‰1).Â It obtains realistic long-wavelength response properties of an isolated layer via efficient calculations using standard solid-state packages to provide the macroscopic dielectric function of a â€œstretchedâ€ solid. Linear growth curve model{---}a random-coefficient model. Mixed-effects logistic regression. Bayesian analysis of change-point problem. Bioequivalence in a crossover trial. Random-effects meta-analysis of clinical trials. Item response theory. Stored results. Methods and formulas. Adaptive MH algorithm. Adaptive MH algorithm for random effects. Gibbs sampling for some likelihood-prior and prior-hyperprior configurations. Likelihood-prior configurations. Common item response theory (IRT) models introduce latent person variables to model the dependence between responses of the same participant. Assuming a distribution for the latent variables, these IRT models are formally equivalent with nonlinear mixed models. It is shown how a variety of IRT models can be formulated as particular instances of nonlinear mixed models. The unifying framework offers the advantage that relations between different IRT models become explicit and that it is rather straightforward to see how existing IRT models can be adapted and extended. The approach is illustrated with a self-report study on anger. Item response theory (IRT) is a set of latent variable techniques specically designed to model the interaction between a participants ability, or latent trait, with item level stimuli (diculty, guessing, etc.) Three main reasons to use IRT: Model a test (parameter estimation, diagnostics, dimensionality checking, etc.) in which focus is on the item/population parameters, Explain variability either in the item properties or persons who were given the test, and Score a test to obtain estimates of the latent trait(s) for individual participants. Item response theory Unidimensional IRT Multidimens...