Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50*, 181–202.

Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 155–180). Thousand Oaks, CA: Sage.

Molenaar, P. C. M., & Nesselroade, J. R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-Toeplitz matrices representing single-subject multivariate time series. *Multivariate Behavioral Research, 33*, 313–342.

Molenaar, P. C. M., & Nesselroade, J. R. (2001). Rotation in the dynamic factor modeling of multivariate stationary time series. *Psychometrika, 66*, 99–107.

Nesselroade, J. R., & Ford, D. H. (1987). Methodological considerations in modeling living systems. In M. E. Ford & D. H. Ford (Eds.), *Humans as self-constructing living systems: Putting the framework to work* (pp. 47–79). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Alternative dynamic factor models for multivariate time-series analyses. In D. M. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Advances and techniques* (pp. 235–265). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223–250). Newbury Park, CA: Sage.

Nesselroade, J. R., & Molenaar, P. C. M. (2003). Quantitative models for developmental processes. In J. Valsiner & K. Connolly (Eds.), *Handbook of developmental psychology* (pp. 622–639). London: Sage.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*, 472–490.

Shoda, Y., Mischel, W., & Wright, J. C. (1993). The role of situational demands and cognitive competencies in behavior organization and personality coherence. *Journal of Personality and Social Psychology, 65*, 1023–1035.

Steyer, R., Ferring, D., & Schmitt, M. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*, 79–98.

Vallacher R. R., & Nowak A. (Eds.). (1994). *Dynamical systems in social psychology*. San Diego, CA: Academic.

Wei, W. W. S. (1990). *Time series analysis: Univariate and multivariate methods*. New York: Addison-Wesley.

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality, 9*, 609–662.

Wood, P., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin, 116*, 166–186.

Yule, G. U. (1927). On a method for investigating periodicities in disturbed series with special reference to Wölfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London, Series A, 224*, 267–298.

Zevon, M., & Tellegen, A. (1982). The structure of mood change: Idiographic/nomothetic analysis. *Journal of Personality and Social Pyschology, 43*, 111–122.

Contemporary Psychometrics
A. Maydeu-Olivares + J J McArdle (eds)
288

14

Multilevel Factor Analysis Models for Continuous and Discrete Data

Harvey Goldstein
University of London

William Browne
University of Nottingham

MULTILEVEL MODELS

The technique of multilevel modeling is now well established and has been incorporated into most standard statistical packages. It is convenient for present purposes to consider educational data that exhibit a hierarchical structure of students nested within classrooms nested within schools—a three-level structure. Other examples are repeated-measures data with occasions nested within subjects, and surveys with people nested within households, both of these being two-level data structures. More-complex data structures such as cross-classifications and multiple-membership structures are extensions for which models have been developed. Goldstein (2003) provided a detailed exposition with references to further application areas.

A general model for the three-level schooling case, assuming normality, can be written as

$$
\begin{aligned}
y_{ijk} &= (X\beta)_{ijk} + Z^{(3)}_{ijk} w_k + Z^{(2)}_{ijk} v_{jk} + e_{ijk}, \\
w &\sim \text{MVN}(0, \Omega_w), \qquad v \sim \text{MVN}(0, \Omega_v), \qquad (1)
\end{aligned}
$$

where the superscripts indicate the level; the more general n-level model can be written compactly as

$$Y = X\beta + Zu + e, \qquad u \sim \mathrm{MVN}(0, \Omega_u), \qquad e \sim N\left(0, \sigma_e^2 I\right).$$

For a generalized linear model, say with a binary response, we correspondingly have

$$g(\pi) = X\beta + Zu, \qquad u \sim \mathrm{MVN}(0, \Omega_u), \qquad Y \sim \mathrm{Bernoulli}\,(1, \pi), \tag{2}$$

where g is a suitable link function and $X\beta$ refers to the fixed-coefficient (regression) component of the model. We shall assume that the Level 1 residual matrix is diagonal, and this also will apply to our factor models.

First, we review briefly the traditional approach to estimating these models based on maximum likelihood.

MAXIMUM LIKELIHOOD ESTIMATION

For the normal model the standard (twice) the log likelihood is

$$\begin{aligned} 2L(\Omega, \beta) &= -\log|V| - \mathrm{tr}(V^{-1}S) = -\log|V| - (Y - X\beta)^T V^{-1}(Y - X\beta), \\ S &= (Y - X\beta)(Y - X\beta)^T, \qquad V = \mathrm{cov}(Y|X\beta), \end{aligned} \tag{3}$$

where Ω is the set of random parameters comprising the variances and covariances in Equation 1. If we have an ML estimate of β, then

$$2L(\Omega, \hat\beta) = -\log|V| - \mathrm{tr}(V^{-1}S) = -\log|V| - (Y - X\hat\beta)^T V^{-1}(Y - X\hat\beta) \tag{4}$$

is the profile likelihood for the random parameters Ω. A convenient algorithm known as iterative generalized least squares (IGLS) alternates between maximizing Equation 4 and then obtaining the conditional ML (GLS) estimate of Ω until convergence.

We can write the extended likelihood, referred to in different contexts as a penalized likelihood or an h-likelihood (Lee & Nelder, 2001), that includes the actual random effects (residuals) as parameters:

$$\begin{aligned} 2L(\Omega, \beta, u) &= -\log|R| - (Y - X\beta - Zu)^T R^{-1}(Y - X\beta - Zu) \\ &\quad -\log|\Omega_u| - u^T \Omega_u^{-1} u, \\ R &= \sigma_e^2 I. \end{aligned} \tag{5}$$

If we maximize Equation 5 for the random effects, given (β, Ω), we obtain the usual estimator, which can be written conveniently as

$$\hat u = \left(Z^T R^{-1} Z + \Omega_u^{-1}\right)^{-1} Z^T(Y - X\beta). \tag{6}$$

Given Ω and u, the profile likelihood for the fixed effects is thus

$$\begin{aligned} 2L(\hat\Omega, \hat u) &= -\log|\hat R| - (Y - X\beta - Z\hat u)^T \hat R^{-1}(Y - X\beta - Z\hat u) \\ &\quad -\log|\hat\Omega_u| - \hat u^T \hat\Omega_u^{-1} \hat u. \end{aligned} \tag{7}$$

Thus, a convenient modification of the IGLS procedure is to iterate between calculating the fixed effects using Equation 7, which, when R is diagonal, is just ordinary least squares (OLS), calculating the random effects from Equation 6 and then the random parameters using the same step as is used in the standard IGLS algorithm.

Expression 3 is known as the *marginal* log likelihood because it is obtained by integrating out the random effects regarded as nuisance parameters.

Because the random effects depend on the random but not the fixed parameters, more generally we can write

$$\log[L(\beta, \Omega, U)] = \sum \{\log[f(Y|U;\beta)] + \log[f(U;\Omega)]\}. \tag{8}$$

The marginal likelihood is thus given by

$$L(\beta, \Omega) = \int f(Y|U;\beta) f(U;\Omega)\, dU, \tag{9}$$

where the first term on the right-hand side is the distribution function for the responses conditional on the random effects, or residuals, U. The second term is the distribution function for the random effects. The first term, given U, depends only on the unknown parameters β, and the second depends only on the unknown parameters Ω. Thus, for example, for a two-level logistic binary response model where the random effects are assumed to be multivariate normal we have, because the random effects are independent across units,

$$\begin{aligned} L(\beta, \Omega) &= \prod_j \int \prod_i \left[(\pi_{ij})^{s_{ij}} (1 - \pi_{ij})^{n_{ij} - s_{ij}}\right] \Phi(u_j;\Omega)\, du_j, \\ \pi_{ij} &= [1 + \exp(-X_{ij}\beta_j)]^{-1}, \qquad \beta_j = \beta + u_j, \end{aligned} \tag{10}$$

where Φ is the multivariate normal density function for the u_j, and n_{ij} and s_{ij} are the number of trials and number of successes, respectively. The integral in Equation 10 can be written in the form $\int_{-\infty}^{\infty} P(u_j)\Phi(u_j)\, du_j$.

Gauss–Hermite quadrature approximates such an integral as

$$\int_{-\infty}^{\infty} P(v) e^{-v^2} dv \approx \sum_{q=1}^{Q} P(x_q) w_q, \tag{11}$$

where the right-hand side is a Gauss–Hermite polynomial evaluated at a series of quadrature points indexed by q. Hedeker and Gibbons (1994) gave a detailed discussion and also considered the multicategory (multinomial) response case. This function is then maximized using a suitable search procedure over the parameter space. Rabe-Hesketh, Pickles, and Taylor (2002) used quadrature to fit general multilevel structural equation models with a variety of link functions. An alternative to quadrature is to use simulated maximum likelihood, which is attractive for models with large numbers of random parameters (Goldstein, 2003, Appendix 4.2).

We now look at multilevel factor models. We briefly refer to the maximum likelihood analysis of multilevel factor analysis models and then develop an alternative approach using Markov chain Monte Carlo (MCMC) estimation.

A MULTILEVEL FACTOR MODEL

We begin by considering a simple single-level factor model for continuous responses, which we write as

$$\begin{aligned} y_{ri} &= \lambda_r v_i + e_{ri}, \qquad r = 1, \ldots, R, \qquad i = 1, \ldots, N, \\ v_i &\sim N(0, 1), \qquad e_{ri} \sim N\left(0, \sigma_{er}^2\right), \end{aligned} \tag{12}$$

where r indexes the responses and i indexes individuals. This can in fact be viewed as a two-level model with a single Level 2 random effect (v_i) with variance constrained to 1 and R Level 1 units for each Level 2 unit, each with its own (unique) variance.

If we knew the values of the "loadings" λ_r, then we could fit Equation 12 directly as a two-level model with the loading vector as the explanatory variable for the Level 2 variance, which is constrained to be equal to 1; if there are any measured covariates in the model, their coefficients can be estimated at the same time. Conversely, if we knew the values of the random effects v_i, we could estimate the loadings; this would now be a single-level model with each response variate having its own variance. These considerations suggest that an expectation-maximization algorithm can be used in the estimation where the random effects are regarded as missing data (Rubin & Thayer, 1982). They also motivate the use of MCMC estimation, which we discuss later.

We now add a second level with its own factor structure and write

$$\begin{aligned} Y &= \Lambda_2 v_2 + u + \Lambda_1 v_1 + e, \\ Y &= \{y_{rij}\}, \qquad u = \{u_r\}, \qquad e = \{e_r\}, \\ r &= 1, \ldots, p, \qquad i = 1, \ldots, n_j, \qquad j = 1, \ldots, J, \\ v_2 &\sim N(0, 1), \qquad v_1 \sim N(0, 1), \qquad e_r \sim N\left(0, \sigma_{er}^2\right), \qquad u_r \sim N\left(0, \sigma_{ur}^2\right) \end{aligned} \tag{13}$$

where the "uniquenesses" u (Level 2) and e (Level 1) are mutually independent and there are p response measures. Here Λ_1 and Λ_2 are the loading matrices for the Level 1 and Level 2 factors, respectively, and v_1 and v_2 are the independent factor vectors at Level 1 and Level 2, respectively. Note that we can have different numbers of factors at each level. We adopt the convention of regarding the measurements themselves as constituting the lowest level of the hierarchy, so that Equation 13 is regarded as a three-level model. Extensions to more levels are straightforward.

We can write the normal log likelihood for Equation 12 as

$$\begin{aligned} 2L(\Omega, \beta, \nu) &= -\log|R| - (Y - \Lambda\nu)^T R^{-1}(Y - \Lambda\nu) \\ &\quad -\log|\Omega_\nu| - \nu^T \Omega_\nu^{-1} \nu, \\ \Lambda &= (\Lambda_1\,\Lambda_2), \qquad \nu^T = (\nu_1\,\nu_2), \qquad R = \operatorname{diag}\left(\sigma_{el}^2, \ldots, \sigma_{ep}^2\right), \end{aligned} \tag{14}$$

with corresponding expressions for other link functions. A general approach to estimation is to form the marginal likelihood as described earlier. McDonald and Goldstein (1989) provided an explicit computational algorithm for the normal response model of Equation 13. Longford and Muthen (1992) developed this approach. The latter authors, together with Goldstein (2003, Chapter 7) and Rowe and Hill (1997) and Rowe (2003), also pointed out that consistent estimators can be obtained from a two-stage process as follows. A two-level multivariate response linear model is fitted using an efficient procedure such as maximum likelihood. This can be accomplished, for example, as pointed out earlier by defining a three-level model where the lowest level is that of the response variables (Goldstein, 2003, Chapter 6; also see the model in Equation 15). This analysis will produce estimates for the (residual) covariance matrices at each level, and each of these can then be structured according to an underlying latent variable model in the usual way. By considering the two matrices as two "populations," we can also impose constraints on, say, the loadings using an algorithm for simultaneously fitting structural equations across several populations.

This chapter describes a general approach to the estimation of such multilevel factor analysis models using MCMC. In the standard multilevel model described in Equation 1, MCMC treats the random effects at higher levels as parameters alongside the fixed coefficients, variances, and covariances. The algorithm proceeds in steps where at each step a parameter or set of parameters is updated by sampling from the distribution for those parameters conditional on the current values for all the other parameters, the data, and the prior distributions. For each parameter this results in a chain of correlated values, which, after the chain has become stationary, can be used for inference. Thus, the mean and the standard deviation provide estimates corresponding to the traditional maximum likelihood estimate and its standard error. An advantage of MCMC is that, given a long enough chain, we can obtain exact interval estimates based on quantiles rather than relying on large-sample approximations.

We now describe the details of an MCMC algorithm for the factor analysis model.

MARKOV CHAIN MONTE CARLO ESTIMATION FOR THE FACTOR ANALYSIS MODEL

We first develop our MCMC algorithm for the multivariate normal model. This is followed by an extension to the binary and mixed response cases, where we also give a detailed example. Further discussion of the multivariate normal model can be found in Goldstein and Browne (2002).

To show the steps of the MCMC algorithm we write Equation 13 in the more detailed form

$$y_{rij} = \beta_r + \sum_{f=1}^{F} \lambda_{fr}^{(2)} \nu_{fj}^{(2)} + \sum_{g=1}^{G} \lambda_{gr}^{(1)} \nu_{gij}^{(1)} + u_{rj} + e_{rij}, \qquad u_{rj} \sim N\left(0, \sigma_{ur}^2\right),$$

$$e_{rij} \sim N\left(0, \sigma_{er}^2\right), \qquad \nu_{fj}^{(2)} \sim \mathrm{MVN}_F(0, \Omega_2), \qquad \nu_{gij}^{(1)} \sim \mathrm{MVN}_G(0, \Omega_1), \quad (15)$$

$$r = 1, \ldots, R, \qquad i = 1, \ldots, n_j, \qquad j = 1, \ldots, J, \qquad \sum_{j=1}^{J} n_j = N.$$

Again we have R responses for N individuals split between J Level 2 units. We have F sets of factors $\nu_{fj}^{(2)}$ defined at Level 2, and G sets of factors $\nu_{gij}^{(1)}$ defined at Level 1. We also introduce the fixed part of the model, but for simplicity restrict our algorithm to a single intercept term β_r for each response, although it is easy to extend the algorithm to arbitrary fixed terms. The residuals at Levels 1 and 2, e_{rij} and u_{rj}, respectively, are assumed to be independent.

Although this allows a very flexible set of factor models, it should be noted that for such models to be identifiable, suitable constraints must be put on the parameters. See Everitt (1984) for further discussion of identifiability. These consist in fixing the values of some of the elements of the factor variance matrices Ω_1 and Ω_2 and/or some of the factor loadings $\lambda_{fr}^{(2)}$ and $\lambda_{gr}^{(1)}$.

The algorithms presented will give steps for all parameters, and so any parameter that is constrained will simply maintain its chosen value and will not be updated. We initially assume that the factor variance matrices Ω_1 and Ω_2 are known (completely constrained) and then discuss how the algorithm can be extended to encompass partially constrained variance matrices. The parameters in the following steps are those available at the current iteration of the algorithm.

Prior Distributions

For the algorithm we assume the following general priors:

$$p(\beta_r) \sim N\left(\beta_r^*, \sigma_{br}^2\right),$$

$$p(\lambda_{fr}^{(2)}) \sim N(\lambda_{fr}^{(2)*}, \sigma_{2fr}^2), \qquad p\left(\lambda_{gr}^{(1)}\right) \sim N\left(\lambda_{gr}^{(1)*}, \sigma_{1gr}^2\right),$$

$$p\left(\sigma_{ur}^2\right) \sim \Gamma^{-1}\left(a_{ur}^*, b_{ur}^*\right), \qquad p\left(\sigma_{er}^2\right) \sim \Gamma^{-1}\left(a_{er}^*, b_{er}^*\right).$$

Known Factor Variance Matrices

We assume that the factor variance matrices are known, so that we can use a Gibbs sampling algorithm, which involves updating parameters in turn by generating new values from the following eight sets of conditional posterior distributions.

Step 1. Update current value of β_r ($r = 1, \ldots, R$) from the following distribution:

$$p(\beta_r) \sim N\left(D_{br}\left(\frac{\sum_{ij} d_{rij}^{\beta}}{\sigma_{er}^2} + \frac{\beta_r^*}{\sigma_{br}^2}\right), D_{br}\right),$$

where

$$D_{br} = \left(\frac{N}{\sigma_{er}^2} + \frac{1}{\sigma_{br}^2}\right)^{-1} \quad \text{and} \quad d_{rij}^{\beta} = e_{rij} + \beta_r.$$

Step 2. Update $\lambda_{fr}^{(2)}$ ($r = 1, \ldots, R; f = 1, \ldots, F$, where not constrained) from the following distribution:

$$p(\lambda_{fr}^{(2)}) \sim N\left(D_{fr}^{(2)}\left(\frac{\sum_{ij} \nu_{fj}^{(2)} d_{rijf}^{(2)}}{\sigma_{er}^2} + \frac{\lambda_{fr}^{(2)*}}{\sigma_{2fr}^2}\right), D_{fr}^{(2)}\right),$$

where

$$D_{fr}^{(2)} = \left(\frac{\sum_j n_j \left(\nu_{fj}^{(2)}\right)^2}{\sigma_{er}^2} + \frac{1}{\sigma_{2fr}^2}\right)^{-1} \quad \text{and} \quad d_{rijf}^{(2)} = e_{rij} + \lambda_{fr}^{(2)} \nu_{fj}^{(2)}.$$

Step 3. Update $\lambda_{gr}^{(1)}$ ($r = 1, \ldots, R; g = 1, \ldots, G$, where not constrained) from the following distribution:

$$p\left(\lambda_{gr}^{(1)}\right) \sim N\left(D_{gr}^{(1)}\left(\frac{\sum_{ij} \nu_{gij}^{(1)} d_{rijg}^{(1)}}{\sigma_{er}^2} + \frac{\lambda_{gr}^{(1)*}}{\sigma_{1gr}^2}\right), D_{gr}^{(1)}\right),$$

where

$$D_{gr}^{(1)} = \left(\frac{\sum_{ij} \left(v_{gij}^{(1)}\right)^2}{\sigma_{er}^2} + \frac{1}{\sigma_{1gr}^2} \right)^{-1} \quad \text{and} \quad d_{rijg}^{(1)} = e_{rij} + \lambda_{gr}^{(1)} v_{gij}^{(1)}.$$

Step 4. Update $v_j^{(2)}$ $(j = 1, \ldots, J)$ from the following distribution:

$$p\left(v_j^{(2)}\right) \sim \mathrm{MVN}_F \left(D_j^{(2)} \left(\sum_r \sum_{i=1}^{n_j} \frac{\lambda_r^{(2)} d_{rij}^{(2)}}{\sigma_{er}^2} \right), D_j^{(2)} \right),$$

where

$$D_j^{(2)} = \left(\sum_r \frac{n_j \lambda_r^{(2)} \left(\lambda_r^{(2)}\right)^T}{\sigma_{er}^2} + \Omega_2^{-1} \right)^{-1}$$

and

$$d_{rij}^{(2)} = e_{rij} + \sum_{f=1}^{F} \lambda_{fr}^{(2)} v_{fj}^{(2)}, \quad \lambda_r^{(2)} = \left(\lambda_{1r}^{(2)}, \ldots, \lambda_{Fr}^{(2)}\right)^T, \quad v_j^{(2)} = \left(v_{1j}^{(2)}, \ldots, v_{Fj}^{(2)}\right)^T.$$

Step 5. Update $v_{ij}^{(1)}$ $(i = 1, \ldots, n_j; j = 1, \ldots, J)$ from the following distribution:

$$p\left(v_{ij}^{(1)}\right) \sim \mathrm{MVN}_G \left(D_{ij}^{(1)} \left(\sum_r \frac{\lambda_r^{(1)} d_{rij}^{(1)}}{\sigma_{er}^2} \right), D_{ij}^{(1)} \right),$$

where

$$D_{ij}^{(1)} = \left(\sum_r \frac{\lambda_r^{(1)} \left(\lambda_r^{(1)}\right)^T}{\sigma_{er}^2} + \Omega_1^{-1} \right)^{-1}$$

and

$$d_{rij}^{(1)} = e_{rij} + \sum_{g=1}^{G} \lambda_{gr}^{(1)} v_{gj}^{(1)}, \quad \lambda_r^{(1)} = \left(\lambda_{1r}^{(1)}, \ldots, \lambda_{Gr}^{(1)}\right)^T, \quad v_{ij}^{(1)} = \left(v_{1ij}^{(1)}, \ldots, v_{Gij}^{(1)}\right)^T.$$

Step 6. Update u_{rj} $(r = 1, \ldots, R; j = 1, \ldots, J)$ from the following distribution:

$$p(u_{rj}) \sim N \left(\frac{D_{rj}^{(u)}}{\sigma_{er}^2} \sum_{i=1}^{n_j} d_{rij}^{(u)}, D_{rj}^{(u)} \right),$$

where

$$D_{rj}^{(u)} = \left(\frac{n_j}{\sigma_{er}^2} + \frac{1}{\sigma_{ur}^2} \right)^{-1} \quad \text{and} \quad d_{rij}^{(u)} = e_{rij} + u_{rj}.$$

Step 7. Update σ_{ur}^2 from the following distribution:

$$p\left(\sigma_{ur}^2\right) \sim \Gamma^{-1}(\hat{a}_{ur}, \hat{b}_{ur}),$$

where

$$\hat{a}_{ur} = J/2 + a_{ur}^* \quad \text{and} \quad \hat{b}_{ur} = \frac{1}{2} \sum_j u_{rj}^2 + b_{ur}^*.$$

Step 8. Update σ_{er}^2 from the following distribution:

$$p\left(\sigma_{er}^2\right) \sim \Gamma^{-1}(\hat{a}_{er}, \hat{b}_{er}),$$

where

$$\hat{a}_{er} = N/2 + a_{er}^* \quad \text{and} \quad \hat{b}_{er} = \frac{1}{2} \sum_{ij} e_{rij}^2 + b_{er}^*.$$

Note that the Level 1 residuals e_{rij} can be calculated by subtraction at every step of the algorithm.

Unconstrained Factor Covariances

In the general algorithm we have assumed that the factor variances are all constrained. Typically we will fix the variances to equal 1 and the covariances to equal 0 and have independent factors. This form will allow us to simplify Steps 4 and 5 of the algorithm to univariate normal updates for each factor separately. However, we may wish to consider correlations between the factors. Here we modify our algorithm to allow another special case, where the variances are constrained to be 1 but the covariances can be freely estimated. Where the resulting correlations obtained are estimated to be close to 1 or −1, then we may be fitting too many factors at that particular level. As the variances are constrained to equal 1, the covariances between factors equal the correlations between the factors. This means that each covariance is constrained to lie between −1 and 1. We consider here only the factor variance matrix at Level 2 because the step for the Level 1 variance matrix simply involves changing subscripts. We use the following priors:

$$p(\Omega_{2,lm}) \sim \mathrm{Uniform}(-1, 1) \qquad \forall l \neq m.$$

Here $\Omega_{2,lm}$ is the lmth element of the Level 2 factor variance matrix. We update these covariance parameters using a Metropolis step and a normal random walk proposal [see Browne (2003) for more details on using Metropolis–Hastings methods for constrained variance matrices].

Step 9. At iteration t generate $\Omega^*_{2,lm} \sim N(\Omega^{(t-1)}_{2,lm}, \sigma^2_{plm})$, where σ^2_{plm} is a proposed distribution variance that has to be set for each covariance. Then, if $\Omega^*_{2,lm} > 1$ or $\Omega^*_{2,lm} < -1$, set $\Omega^{(t)}_{2,lm} = \Omega^{(t-1)}_{2,lm}$ as the proposed covariance is not valid, else form a proposed new matrix Ω^*_2 by replacing the lmth element of $\Omega^{(t-1)}_2$ by this proposed value. We then set

$$\Omega^{(t)}_{2,lm} = \Omega^*_{2,lm} \quad \text{with probability} \quad \min\left(1, p(\Omega^*_2|\nu^{(2)}_j)/p(\Omega^{(t-1)}_2|\nu^{(2)}_j)\right)$$
$$\Omega^{(t)}_{2,lm} = \Omega^{(t-1)}_{2,lm} \quad \text{otherwise.}$$

Here

$$p(\Omega^*_2|\nu^{(2)}_j) = \prod_j |\Omega^*_2|^{-1/2} \exp\left[((\nu^{(2)}_j)^T(\Omega^*_2)^{-1}\nu^{(2)}_j)/2\right]$$
$$p(\Omega^{(t-1)}_2|\nu^{(2)}_j) = \prod_j |\Omega^{(t-1)}_2|^{-1/2} \exp\left[((\nu^{(2)}_j)^T(\Omega^{(t-1)}_2)^{-1}\nu^{(2)}_j)/2\right].$$

This procedure is repeated for each covariance that is not constrained.

Missing Data

Where some of the responses are missing, this poses no problem for the MCMC methods if we are prepared to assume missingness is at random or effectively so by design. This is equivalent to giving the missing data a uniform prior. We then have to simply add an extra Gibbs sampling step to the algorithm to sample the missing values at each iteration. As an illustration, we consider an individual who is missing response r. In a factor model the correlation between responses is explained by the factor terms, and conditional on these terms the responses for an individual are independent, and so the conditional distributions of the missing responses have simple forms.

Step 10. Update y_{rij} $(r = 1, \ldots, R; i = 1, \ldots, n_j; j = 1, \ldots, J; \forall y_{rij}$ that are missing) from the following distribution, given the current values $y_{rij} \sim N(\hat{y}_{rij}, \sigma^2_{er})$ where

$$\hat{y}_{rij} = \beta_r + \sum_{f=1}^{F} \lambda^{(2)}_{fr}\nu^{(2)}_{fj} + \sum_{g=1}^{G} \lambda^{(1)}_{gr}\nu^{(1)}_{gij} + u_{rj}.$$

Goldstein and Browne (2002) discussed the extension of this model to the general structural equation case.

BINARY RESPONSE FACTOR MODELS

Modeling of data that consist of binary or ordered responses to questions in an achievement or similar test instrument has a long history. Goldstein and Wood (1989) described the history of mental testing from the early work of Lawley (1943) and the work of Lord and Novick (1968) on item response models to more recent developments of general factor analysis modeling (Bock, Gibbons, & Muraki, 1988). Rod McDonald made important contributions to this area through his discussions of test item dimensionality and models for nonlinear factor analysis (McDonald, 1981, 1985).

The early work was characterized by "fixed-effect" models of the kind

$$f(\pi_{ri}) = \beta_{0r} + \beta_{1r}\theta_i \qquad (16)$$

relating the probability of a correct response to the rth item for the ith respondent, where typically a logit link function is used for the probability. The most common link function f is a logit or probit. The response y is typically (0, 1), and we have the local, or conditional, independence assumption

$$y_{ij} \overset{\text{iid}}{\sim} Bin(1, p_{ij}).$$

This is often referred to, somewhat inaccurately, as a two-parameter model, where in Equation 16 each response is characterized by an intercept β_{0r} and a factor coefficient β_{1r}, and each respondent has a factor value θ_i. This gives rise to a model with $N + 2p$ parameters, where N is the number of respondents and p is the number of items or questions. Extensions to the case where responses are on an ordered scale (a graded response or partial credit model; Baker, 1992) relate the cumulative proportion of success to a linear function via a suitable link function, for example, the cumulative odds model for category h of item r:

$$f\left(\sum_{g=1}^{h} \pi_{gri} \Big/ \sum_{g=h+1}^{t_r} \pi_{gri}\right) = \beta_{0rh} + \beta_{1r}\theta_i, \qquad h = 1, \ldots, t_{r-1}, \qquad (17)$$

where t_r indexes the final category of item r.

Such fixed-effect models have more recently been superseded by "random effects" models (Bartholomew & Knott, 1999), where the individual parameter θ_i is assumed to have a distribution, typically normal, across individuals. This provides both more efficient estimates and straightforward ways of handling missing responses. More important, it allows for the fitting of more than one parameter for individuals, so that we can write down a general multidimensional binary (or ordered) extension of Equation 16,

$$f(\pi_{ri}) = \beta_{0r} + \sum_{h=1}^{q} \beta_{hr}\theta_{hi}, \qquad \theta_i \sim \mathrm{MVN}(0, \Omega),$$
$$y_{ri} \overset{\text{iid}}{\sim} Bin(1, \pi_{ri}), \tag{18}$$

which is simply a single-level binary response factor model.

Having fitted such a model, we can obtain estimates of factor values, or scores, for each individual on each factor. In practice, models with more than one factor dimension have been used rarely, or typically are covariates incorporated, for example, for gender or other predictors. We explore some of the consequences of this in the analysis of a large-scale data set.

We now introduce a multilevel model, as with the normal response case, that recognizes that groups such as schools may differ in their response probabilities. We write

$$f(\pi_{rij}) = \beta_{0r} + \sum_{h=1}^{q_1} \beta_{hr}^{(1)}\theta_{hij}^{(1)} + \sum_{h=1}^{q_2} \beta_{hr}^{(2)}\theta_{hj}^{(2)} + u_{rj},$$
$$\theta_{ij}^{(1)} \sim \mathrm{MVN}(0, \Omega_1), \qquad \theta_j^{(2)} \sim \mathrm{MVN}(0, \Omega_2), \tag{19}$$
$$y_{rjk} \overset{\text{iid}}{\sim} Bin(1, \pi_{rjk}).$$

We have now added a second set of factors, indexed by the superscript (2), varying at the group Level 2, independent of the individual level factors, indexed by the superscript (1). In contrast to the normal response factor model, the Level 1 variance is constrained by the assumption of binomial variation, and the factor structure has a nonlinear link with the responses. We retain the notational conventions for binary response models, generalized from Equation 17, where we have the following equivalences for the factor structure between Equations 18 and 15:

$$\beta_{hr}^{(1)} \equiv \lambda_{gr}^{(1)}, \qquad \beta_{hr}^{(2)} \equiv \lambda_{fr}^{(2)},$$
$$\theta_{hij}^{(1)} \equiv v_{gij}^{(1)}, \qquad \theta_{hj}^{(2)} \equiv v_{fj}^{(2)}.$$

The Level 2 residuals u_{rj} are assumed independent, $N(0, \sigma_{ur}^2)$.

We show how to specify and fit such a model and use it with a large-scale survey of student achievement.

DATA

The data are taken from the Programme for International Student Assessment (PISA) carried out under the auspices of the Organisation for Economic Co-operation and Development (OECD) in 2000 in 32 industrialized countries

(Organisation for Economic Co-operation and Development, 1999). The data sets, together with full descriptions, are available online (www.pisa.oecd.org). The full data set included a student and school questionnaire together with tests of reading, mathematics, and science. A sample of 14- to 15-year-old school students was selected in each country with a 70% response rate as a minimum requirement for inclusion. The OECD program plans further surveys every 3 years. The major aim of the survey was to develop a "literacy scale" for each of the three areas tested and to compare countries in terms of their average performance on these. The resulting continuous scales had distributions approximately normal with a mean of 500 and a standard deviation 100. Each scale was divided into six "proficiency" categories, each characterized in terms of responses to chosen sample test items. The scores were also analyzed by background factors such as gender and parental education, and a multilevel analysis was also carried out to study variation between schools.

The three OECD literacy scales were constructed using the model of Equation 18 with a single factor. Each scale used only those items designated as Reading, Mathematics, or Science. Factor scores were computed for use in subsequent analyses. For these analyses a multiple imputation procedure was used as follows.

Each student has a factor score based on a linear predictor using his or her individual responses and the estimated model parameters. Under the model assumptions, these scores have an approximate normal distribution, the accuracy of the approximation being a function of the number of item responses for an individual. Using an estimate of the standard error, multiple imputation is used; that is, a set of (typically five) random draws from this estimated normal distribution is made and these are then used for subsequent modeling (Rubin, 1996).

We refer to some of the limitations of this as a general procedure later, but for now note that a key feature is the use of a one-dimensional factor model, and we will discuss the incorporation of further dimensions.

For present purposes we chose the Mathematics test items for two countries, France and England. In total there are 31 Math questions. In fact several questions are grouped in that they all relate to the same problem. For example, one problem described a pattern of trees planted as a set of squares of different sizes, and associated with this problem there were three separate questions. For a model such as that of Equation 18 it is doubtful whether for such questions the local independence assumption will hold, although this was assumed in the OECD analysis. A more satisfactory treatment would be to combine the three separate questions into an ordered scale, for example, by forming an a priori suitably weighted combination of the responses, and treating this as an ordered categorical response as described earlier. For present purposes we selected 15 items, each of which is a response to a different problem, and dichotomized responses into correct/incorrect, treating part-correct answers as correct.

ESTIMATION FOR THE BINARY RESPONSE FACTOR MODEL

The OECD analyses use the logit link function. The probit function generally produces similar results and has certain advantages in terms of computational convenience. One important advantage of the probit is that we can think of the response as a threshold from an underlying (unknown) continuous response, which is normally distributed (Albert & Chib, 1993). We use the Gibbs sampling algorithm for normally distributed responses described earlier and adapted for a probit model as follows.

Assume that we have a binary variable y_i collected for several individuals i, that is, a threshold version of an (unknown) continuous normally distributed variable y_i^*. Now, if we knew the value of y_i^*, then we could fit the standard Gibbs sampling algorithm for normal response models. So we add an extra step to the Gibbs sampling algorithm and generate y_i^* at each iteration from its conditional posterior distribution, which is a truncated normal distribution with mean (in the standard single-level probit model) $X\beta$ and variance one. The truncation point is zero, and if y_i is zero, y_i^* has to be negative, and if y_i is one, y_i^* has to be positive. This step is inserted into the existing algorithm for the normal response factor model. It should be noted that this model can also be updated using Metropolis sampling, but the Gibbs sampling algorithm is faster and produces fewer correlated chains. Consider the standard two-level model

$$Y = X\beta + ZU + e, \qquad e \sim N(0, 1). \tag{20}$$

Given current estimates of parameters and residuals, we have $Y \sim N(XB + ZU, 1)$, and for the probit model the observation of a positive value (>0) on the scale of Y corresponds to the observation of a "success" on the probability scale and the observation of a negative (<0) value corresponds to a "failure." The probit function that determines the underlying chance of a correct response is the cumulative probability given by

$$\int_0^\infty \phi(t)\,dt, \qquad \phi(t) \text{ is pdf of } N(X\beta + ZU, 1),$$

or equivalently

$$\int_{-(X\beta+ZU)}^\infty \phi(t)\,dt, \qquad \phi(t) \text{ is pdf of } N(0, 1). \tag{21}$$

Alternatively, if we write the value of the ijth response as

$$y_{ij} = (X\beta)_{ij} + (ZU)_{ij} + e_{ij},$$

a positive value occurs when $y_{ij} > 0$. We then have

$$\Pr(y_{ij} > 0) = \Pr(e_{ij} > -[(X\beta)_{ij} + (ZU)_{ij}]), \tag{22}$$

which leads to Equation 21.

Thus, given current values of β, U, and the observation for a Level 1 unit (0 or 1), we take a random draw e^*. If we observe a 1, then we draw from the truncated normal distribution $[-X^*, \infty]$, $X^* = (X\beta + ZU)$, and if we observe a 0, we sample from $[-\infty, -X^*]$. This is then applied to Equation 20 to give a new value Y^*. This procedure is applied as an extra step in the factor analysis model, for example, before Step 1, with the remaining steps as before.

This approach is readily extended to the case of ordered categories, which can be applied to "partial-credit" models. We assume that there is an underlying normally distributed response and that for p-category observed responses there are $p - 1$ thresholds. Assume a proportional odds model, where for the sth cumulative probability we have (Goldstein, 2003)

$$\text{probit}(\gamma^{(s)}) = \alpha^{(s)} + (X\beta) + ZU,$$

so that corresponding to Equation 21 this gives

$$\int_{-(\alpha^{(s)}+X\beta+ZU)}^\infty \phi(t)\,dt,$$

and sampling is conditional, as before, including the current values of the threshold parameters $\alpha^{(s)}$.

RESULTS

The analyses reported here were carried out using MLwiN Beta version 1.2 (Browne, 2004; Rasbash et al., 2000). Table 14.1 shows a basic model in which a simple single-level probit model is fitted for each item allowing for different country means. We see that, of the 10 statistically significant items, France does better on 4 (all free-response items) and worse on 6 (3 free-response and 3 multiple-choice items) than England. The interpretation of the probit function is that it predicts a value from an underlying standard normal distribution with mean zero and standard deviation one. This can be turned into a probability using the Cumulative distribution function of the standard normal distribution. Thus, for example, the French students are, on average, 0.7 standard deviations ahead of the English for item 136Q01 (a free-response Geometry item) but 0.7 standard deviations behind on item 161Q01 (a multiple-choice Geometry item).

We now fit a single factor at each level (the student and the school), with results under Analysis A in Table 14.2. For convenience we present only the estimated factor loadings. At both levels we have a common factor with comparable loadings

TABLE 14.1
Separate Country Analyses With Probit Response Model for Each Item

Item	*England*	*France–England*
Student level		
33q01 (MC)	0.80	−0.06 (0.05)
34q01 (FR)	−0.25	0.03 (0.06)
37q01 (FR)	0.65	−0.11 (0.07)
124q01 (FR)	0.01	−0.18 (0.07)
136q01 (FR)	−0.23	0.69 (0.05)
144q01 (FR)	0.16	0.40 (0.05)
145q01 (FR)	0.65	−0.13 (0.06)
150q01 (FR)	0.78	−0.35 (0.06)
155q01 (FR)	0.54	0.27 (0.06)
159q01 (MC)	0.89	−0.24 (0.06)
161q01 (MC)	0.96	−0.70 (0.06)
179q01 (FR)	−0.11	0.64 (0.06)
192q01 (MC)	−0.28	0.07 (0.06)
266q01 (MC)	−0.75	−0.26 (0.06)
273q01 (MC)	−0.04	0.03 (0.06)

Note. Columns show the English mean and the French–English difference between means. Standard errors are shown in parentheses. There are 10,000 Markov chain Monte Carlo iterations with default priors. The type of item is shown by each item name. MC, Multiple choice; FR, free response.

on each item, although at the student level the multiple-choice items tend to have smaller loadings. The next model fits a different mean for each item for France and England, namely

$$\begin{aligned}
\text{probit}(\pi_{rijg}) &= \beta_{0r} + \delta_g d_r + \beta_{1r}^{(1)}\theta_{1ij}^{(1)} + \beta_{1r}^{(2)}\theta_{1j}^{(2)} + u_{rj}, \\
\theta_{1ij}^{(1)} &\sim N\left(0, \sigma_{(1)}^2\right), \qquad \theta_{1j}^{(2)} \sim N\left(0, \sigma_{(2)}^2\right), \\
y_{rijg} &\sim Bin(1, \pi_{rijg}), \qquad g = 1, 2 \\
\delta_g &= \begin{cases} 0 & \text{if } g = 1, \\ 1 & \text{if } g = 2. \end{cases}
\end{aligned} \tag{23}$$

Note that in Equation 23, g identifies country and takes the value 1 for England and 2 for France, and we actually fit a global mean vector plus a difference term, d_r, which captures the difference between French and English scores. The factor loadings are virtually unchanged. The means for the two countries, however, do differ somewhat for certain items. Thus, given the factor values, the French are somewhat further ahead than before on item 136Q01. This suggests that there may be different factor structures in the two countries, and we return to this later.

If we ignore the interaction between country and item, it is possible (but not otherwise) to use these models for purposes of comparing countries. There are two natural extensions where we allow the factor means to vary between countries

TABLE 14.2
Factor Loadings With a Single Factor at Each Level

		Analysis B	
Item	*Analysis A*	*Loadings*	*France–England*
Student level			
33q01	0.46 (0.04)	0.46 (0.04)	−0.08 (0.04)
34q01	0.70 (0.04)	0.71 (0.05)	0.01 (0.07)
37q01	0.96 (0.09)	0.92 (0.07)	−0.16 (0.10)
124q01	0.69 (0.07)	0.72 (0.07)	−0.20 (0.10)
136q01	0.69 (0.05)	0.70 (0.06)	0.96 (0.08)
144q01	0.55 (0.05)	0.54 (0.05)	0.46 (0.07)
145q01	0.63 (0.05)	0.62 (0.05)	−0.21 (0.08)
150q01	0.59 (0.05)	0.59 (0.04)	−0.41 (0.07)
155q01	0.51 (0.05)	0.52 (0.04)	0.33 (0.07)
159q01	0.46 (0.04)	0.47 (0.05)	−0.31 (0.07)
161q01	0.30 (0.04)	0.33 (0.04)	−0.78 (0.07)
179q01	0.54 (0.06)	0.52 (0.06)	0.79 (0.08)
192q01	0.68 (0.04)	0.68 (0.05)	0.09 (0.07)
266q01	0.36 (0.05)	0.38 (0.05)	−0.28 (0.07)
273q01	0.47 (0.04)	0.46 (0.05)	0.03 (0.06)
School level			
33q01	0.26 (0.03)	0.26 (0.03)	
34q01	0.39 (0.03)	0.39 (0.03)	
37q01	0.77 (0.06)	0.74 (0.06)	
124q01	0.71 (0.05)	0.71 (0.05)	
136q01	0.49 (0.04)	0.54 (0.04)	
144q01	0.31 (0.04)	0.31 (0.03)	
145q01	0.47 (0.04)	0.48 (0.04)	
150q01	0.41 (0.04)	0.42 (0.04)	
155q01	0.29 (0.04)	0.31 (0.03)	
159q01	0.35 (0.04)	0.36 (0.04)	
161q01	0.23 (0.04)	0.24 (0.03)	
179q01	0.42 (0.05)	0.46 (0.04)	
192q01	0.43 (0.03)	0.43 (0.04)	
266q01	0.33 (0.04)	0.33 (0.04)	
273q01	0.35 (0.03)	0.35 (0.03)	

Note. Analysis A ignores country differences; analysis B fits the model of Equation 23 and shows the loadings together with French–English difference. Factor variances are set equal to 1.

but where the factor structures are the same in each country. Thus we can extend Equation 23 for country g as follows:

$$\begin{aligned}
\text{probit}(\pi_{rijg}) &= \beta_{0rg} + \beta_{1r}^{(1)}\theta_{1ijg}^{(1)} + \beta_{1r}^{(2)}\theta_{1jg}^{(2)} + u_{rj}, \\
\theta_{1ijg}^{(1)} &\sim N\left(\mu_g^{(1)}, \sigma_{(1)}^2\right), \qquad \theta_{1jg}^{(2)} \sim N\left(\mu_g^{(2)}, \sigma_{(2)}^2\right) \\
y_{rijg} &\sim Bin(1, \pi_{rijg}), \qquad g = 1, 2.
\end{aligned} \tag{24}$$

Typically we would be interested in modeling the same overall shift at each level l, so that we have $\mu_g^{(l)} = \mu_g$. In this case for a single-factor model, we can write Equation 24 in the alternative form

$$\begin{aligned}\text{probit}(\pi_{rijg}) &= \beta_{0r} + \delta_g d\left(\beta_{1r}^{(1)} + \beta_{1r}^{(2)}\right) + \beta_{1r}^{(1)}\theta_{1ij}^{(1)} + \beta_{1r}^{(2)}\theta_{1j}^{(2)} + u_{rj},\\ \theta_{1ij}^{(1)} &\sim N\left(0, \sigma_{(1)}^2\right), \qquad \theta_{1j}^{(2)} \sim N\left(0, \sigma_{(2)}^2\right) \qquad (25)\\ y_{rijg} &\sim Bin(1, \pi_{rijg}), \qquad g = 1, 2.\end{aligned}$$

Clearly we can extend such a model to other explanatory variables such as gender, in effect obtaining a structural equation model for the factor mean structures. We note that the OECD procedure for country and group comparisons will not in general produce the same inferences because the model that is fitted assumes no differences. Likewise, the OECD model assumes only a single (student) level, with school-level variation estimated in the second-stage analysis. In the case of factor models the OECD approach to group comparisons leads to interpretational difficulties because the factor structure that is fitted, in the presence of real group differences under a model such as Equation 23, is incorrect. We also note that for those with at least one valid mathematics item response (55%), the average number of Mathematics items responded to by students is 12.6, with a range from 1 to 16, so that the normal approximation implicit in the use of plausible values may not be very accurate for some of the students.

The OECD country comparisons for Mathematics show a small difference between England and France, and this is borne out by our results, although we have used a reduced set of items.

An alternative formulation for country and group differences is to write Equation 25 as

$$\begin{aligned}\text{probit}(\pi_{rij}) &= \beta_{0r} + d\delta_g + \beta_{1r}^{(1)}\theta_{1ij}^{(1)} + \beta_{1r}^{(2)}\theta_{1j}^{(2)} + u_{rj},\\ \theta_{1ij}^{(1)} &\sim N\left(0, \sigma_{(1)}^2\right), \qquad \theta_{1j}^{(2)} \sim \text{MVN}\left(0, \sigma_{(2)}^2\right), \qquad (26)\\ y_{rjk} &\sim Bin(1, \pi_{rjk}).\end{aligned}$$

This model additionally constrains the item differences for each country to be constant. If we estimate the parameter d in Equation 26, we obtain the value 0.02 with standard error 0.05, so that we would conclude that the French–English difference is small and nonsignificant. In Table 14.2, however, we show considerable individual differences. Thus, if we had only fitted Equation 26 representing a simple overall country effect, as in the PISA analysis, we would be missing potentially important differential (interaction) effects.

The next model fits two orthogonal factors at the student level and one at school level. In the present set of analyses we do not report fitting more than one factor at the school level. In Table 14.3 the first factor at the student level is again a general

TABLE 14.3
Loadings for Two Orthogonal Factors at Level 1 and One Factor at Level 2

Item	*Factor 1*	*Factor 2*
Student level		
33q01	0.51 (0.06)	0
34q01	0.67 (0.05)	0.22 (0.09)
37q01	0.81 (0.10)	0.42 (0.14)
124q01	0.56 (0.11)	0.80 (0.21)
136q01	0.60 (0.09)	0.47 (0.12)
144q01	0.58 (0.10)	0.08 (0.10)
145q01	0.57 (0.06)	0.19 (0.12)
150q01	0.72 (0.10)	−0.07 (0.18)
155q01	0.44 (0.06)	0.28 (0.10)
159q01	0.50 (0.06)	−0.04 (0.12)
161q01	0.43 (0.07)	−0.27 (0.14)
179q01	0.46 (0.08)	0.46 (0.17)
192q01	0.62 (0.06)	0.28 (0.10)
266q01	0.41 (0.06)	−0.10 (0.09)
273q01	0.42 (0.06)	0.21 (0.12)
School level		
33q01	0.27 (0.03)	
34q01	0.39 (0.04)	
37q01	0.76 (0.06)	
124q01	0.82 (0.10)	
136q01	0.52 (0.05)	
144q01	0.32 (0.04)	
145q01	0.47 (0.04)	
150q01	0.45 (0.05)	
155q01	0.31 (0.04)	
159q01	0.36 (0.04)	
161q01	0.25 (0.04)	
179q01	0.46 (0.05)	
192q01	0.44 (0.04)	
266q01	0.34 (0.04)	
273q01	0.36 (0.03)	

Note. The first loading of factor 2 is constrained to zero. The variances constrained to one.

common factor, and the second factor tends to distinguish the free-response from the multiple-choice items. We also studied three factors at the student level, but the results are not easy to interpret, perhaps unsurprisingly given only 15 binary response variables.

We now fit separate factors for the two countries. Table 14.4 shows the results for a single factor at each level. We see that there are different patterns of loadings at both levels and those for France are much closer to the factor loadings estimated from the combined country data set, perhaps unsurprisingly because there are almost twice as many French students in the combined sample. We computed the

TABLE 14.4
Loadings for Single-Factor Models Separately for Each Country

Item	*England*	*France*
Student level		
33q01	0.49 (0.08)	0.46 (0.04)
34q01	0.56 (0.10)	0.75 (0.05)
37q01	0.69 (0.12)	1.09 (0.11)
124q01	0.50 (0.10)	0.82 (0.09)
136q01	0.75 (0.14)	0.71 (0.06)
144q01	0.48 (0.09)	0.58 (0.06)
145q01	0.38 (0.09)	0.68 (0.05)
150q01	0.50 (0.11)	0.62 (0.06)
155q01	0.31 (0.08)	0.59 (0.05)
159q01	0.34 (0.09)	0.51 (0.06)
161q01	0.32 (0.09)	0.33 (0.05)
179q01	0.34 (0.09)	0.62 (0.07)
192q01	0.75 (0.14)	0.68 (0.05)
266q01	0.33 (0.09)	0.41 (0.06)
273q01	0.44 (0.09)	0.48 (0.05)
School level		
33q01	0.36 (0.06)	0.22 (0.04)
34q01	0.46 (0.07)	0.36 (0.04)
37q01	0.73 (0.09)	0.79 (0.08)
124q01	0.80 (0.10)	0.68 (0.06)
136q01	0.75 (0.08)	0.44 (0.04)
144q01	0.41 (0.06)	0.25 (0.04)
145q01	0.72 (0.08)	0.39 (0.04)
150q01	0.44 (0.06)	0.43 (0.04)
155q01	0.23 (0.06)	0.35 (0.04)
159q01	0.47 (0.07)	0.31 (0.04)
161q01	0.27 (0.06)	0.24 (0.04)
179q01	0.50 (0.07)	0.45 (0.05)
192q01	0.42 (0.08)	0.44 (0.04)
266q01	0.38 (0.07)	0.30 (0.05)
273q01	0.41 (0.06)	0.32 (0.04)

factor scores for the English students from the combined and separate analyses and these show a high correlation (.98). This reflects the fact that the factor score is effectively a weighted mean of the item responses, and the two sets of loadings are all positive and comparable in size. It is also inflated because the factor scores are "shrunken" estimates with shrinkage a function of the number of items responded to. A simple comparison of the mean factor scores from the joint analysis with a single factor at each level gives a nonsignificant difference. Thus, whereas a joint analysis will lead to comparable rankings for individuals, as indeed will a simple scoring system using just the average percentage correct (the correlation is .84), the interpretation of factor loadings in the context of group differences will not be the same.

At the school level for France the factor loadings are approximately proportional to those at the school level for the combined analysis, but this is not the case for England, which has different orderings for the loadings. In the pooled analysis the comparison between student and school loadings is more like that for France.

CONCLUSIONS FROM THE ANALYSIS

We have not attempted to study reasons for country differences in any detail in these analyses. Our intention has been to show how multilevel binary factor models can be specified with covariates and group differences. We show that for the purposes of comparing countries it is important to fit a model that explicitly includes country effects. In this case we show that, in the simple case where one general factor is fitted at each level, student and school, the factor structures are somewhat different for each country. Thus, a single "pooled" set of factor loadings used for purposes of country comparisons leads to considerable difficulty in interpreting results because, as in this case, the pooled factors will be influenced by the weightings implicit in the numbers of students in each country. Where several countries are pooled as in the OECD PISA analyses, the factors are even more difficult to interpret, as are resulting country differences. Furthermore, and perhaps more important, we showed that, after fitting a single-factor model there are still differences between countries in item response probabilities (Table 14.2). This implies that the choice of items to use will determine the factor loadings, in that if we choose a majority of items that all load highly on a factor, then that factor will tend to dominate the structure. If those items also happen to favor a particular country then that country will tend to have higher factor scores, but this could only be ascertained by carrying out a multidimensional analysis.

DISCUSSION

The issues that surround the specification and interpretation of single-level factor and structural equation models are also present in our multilevel versions. Parameter identification has already been mentioned; with the ability to include prior distributions we can often treat identification problems with more flexibility. In the traditional model, overparametrization requires setting one or more parameters or functions of parameters to known values. In our case we can obtain estimates by imposing informative prior distributions on each of the parameters, which, when combined with the data, will provide the joint posterior distribution. An example is in the estimation of factor correlations where the assumption of a prior in the interval (0, 1) can allow the joint posterior of all the parameters in an "overidentified" model to be estimated.

Another potential advantage of our approach, common to all MCMC procedures, is that we can make exact inferences based upon the Markov chain values.

This will be a particular advantage for small data sets where we may be unwilling to rely on likelihood-based approximations.

Another issue is the boundary, Heywood case. We observed such solutions occurring where sets of loading parameters tend toward zero or a correlation tends toward 1.0. A final important issue that only affects stochastic procedures is the problem of flipping states. This means that there is not a unique solution even in a one-factor problem because the loadings and factor values may all flip their sign to give an equivalent solution. When the number of factors increases there are greater problems because factors may swap over as the chains progress. This means that identifiability is an important consideration when using stochastic techniques.

We can extend the models considered here to mixtures of binary, ordered, and continuous responses. We have separately discussed all three types of responses. They are linked via the threshold probit model, so that at Level 1 we have a set of independent normal variables (uniquenesses), each one arising from a continuous response, a binary response, or an ordered response. At higher levels the random effects are assumed to have a multivariate normal distribution and the MCMC estimation proceeds in a straightforward fashion.

Such an example might arise in a health application where individuals respond to a health questionnaire at the same time as a set of continuous measurements of health status are made. It might also arise in an educational examination where, for example, some responses are multiple-choice binary questions and some are free responses marked on a continuous scale. Another important application is to questionnaires that contain mixtures of ordered rating scales and binary responses.

A major drawback of current implementations of binary factor (item response) models that attempt to account for multilevel data structures is that they fit a multilevel model in two stages: first estimating a single-level model and then fitting a multilevel model using the estimated factor scores, typically using multiple imputation via plausible values. Such a procedure does not allow the exploration of any factor structure at higher levels. We showed that this may be important, especially when comparing groups or countries.

Finally, we note that all of our models can be extended straightforwardly to more-complex data structures involving cross-classifications and multiple-membership structures (Browne, Goldstein, & Rasbash, 2001).

REFERENCES

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association, 88*, 669–679.

Baker, F. (1992). *Item response theory*. New York: Dekker.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London: Arnold.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261–280.

Browne, W. J. (2002). *MCMC estimation in MLwiN*. London: Institute of Education.

Browne, W. J. (2004). *MCMC estimation of 'constrained' variance matrices with applications in multilevel modelling*. Manuscript submitted for publication.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling, 1*, 103–124.

Everitt, B. S. (1984). *An introduction to latent variable models*. London: Chapman & Hall.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

Goldstein, H., & Browne, W. (2002). Multilevel factor analysis modelling using Markov chain Monte Carlo estimation. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 225–243). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Goldstein, H., & Wood, R. (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology, 42*, 139–167.

Hedeker, D., & Gibbons, R. D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics, 50*, 933–944.

Lawley, D. N. (1943). The application of the maximum likelihood method to factor analysis. *British Journal of Psychology, 33*, 172–75.

Lee, Y., & Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika, 88*, 987–1006.

Longford, N., & Muthen, B. O. (1992). Factor analysis for clustered observations. *Psychometrika, 57*, 581–597.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of mathematical and statistical psychology, 34*, 100–117.

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

McDonald, R. P. (1993). A general model for two-level data with responses missing at random. *Psychometrika, 58*, 575–585.

McDonald, R. P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology, 42*, 215–232.

Meng, X. L. (1994). Posterior predictive *p*-values. *Annals of Statistics, 22*, 1142–1160.

Organisation for Economic Co-operation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.

Rabe-hesketh, S., Pickles, A., & Taylor, C. (2000). Sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin, 53*, 47–57.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., & Lewis T. (2000). *A user's guide to MLwiN* (2nd ed.). London: Institute of Education.

Rowe, K. J., & Hill, P. W. (1997). *Simultaneous estimation of multilevel structural equations to model students' educational progress*. Paper presented at the Tenth International Congress for School Effectiveness and School Improvement, Memphis, TN.

Rowe, K. J. (2003). Estimating interdependent effects among multilevel composite variables in psychosocial research: An example of the application of multilevel structural equation modeling. In S. P. Reise & N. Duan (eds.) *Multilevel modeling: Methodological advances, issues, and applications* (pp. 255–284). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91*, 473–489.

Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika, 47*, 69–76.

Spiegelhalter, D., Best, N., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B, 64*, 583–640.

Zhu, H.-T., & Lee, S.-Y. (1999). Statistical analysis of nonlinear factor analysis models. *British Journal of Mathematical and Statistical Psychology, 52*, 225–242.