British Journal of Mathematical and Statistical Psychology (2010), 63, 293–317

The British Psychological Society
www.bpsjournals.co.uk

Testing and modelling non-normality within the one-factor model

Dylan Molenaar[1]*, Conor V. Dolan[1] and Norman D. Verhelst[2]
[1]University of Amsterdam, The Netherlands
[2]Cito, National Institute for Educational Measurement, MG Arnhem, The Netherlands

Maximum likelihood estimation in the one-factor model is based on the assumption of multivariate normality for the observed data. This general distributional assumption implies three specific assumptions for the parameters in the one-factor model: the common factor has a normal distribution; the residuals are homoscedastic; and the factor loadings do not vary across the common factor scale. When any of these assumptions is violated, non-normality arises in the observed data. In this paper, a model is presented based on marginal maximum likelihood to enable explicit tests of these assumptions. In addition, the model is suitable to incorporate the detected violations, to enable statistical modelling of these effects. Two simulation studies are reported in which the viability of the model is investigated. Finally, the model is applied to IQ data to demonstrate its practical utility as a means to investigate ability differentiation.

1. Introduction

Over the past 100 years, factor analysis has become a widely used technique to infer continuously distributed unobserved variables (common factors) from a larger number of manifest variables. Revealing the factor structure in a set of variables has led to valuable information about psychological constructs related to intelligence (e.g. Dolan, 2000; Johnson & Bouchard, 2004) and personality (e.g. Digman, 1990). The one-factor model is an important case in factor analysis due to its application as an item response model for continuous observations (Mellenbergh, 1994). The one-factor model is formulated as

$$y_{ij} = \upsilon_j + \lambda_j \eta_i + \varepsilon_{ij}, \tag{1}$$

*Correspondence should be addressed to Dylan Molenaar, Psychological Methods, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: D.Molenaar@uva.nl).

DOI:10.1348/000711009X456935

where y_{ij}, the observed value of subject i on variabele j, is regressed on η_i, the common factor score. In this regression, υ_j is the jth intercept, λ_j represents the jth regression slope or factor loading, and ε_{ij} is a residual term. To fit the model from equation (1) to data, various estimation methods have been developed. These methods are often based on the assumption that $\mathbf{y}_i$, the p-dimensional vector of observed variables, is multivariate normally distributed, e.g. generalized least squares (GLS; Bollen, 1989), maximum likelihood (ML; Lawley, 1943) and marginal maximum likelihood (MML; Bock & Aitkin, 1981). Although ML is arguably the most important estimation procedure as it is most frequently used, other methods are available which involve mild distributional assumptions (asymptotically distribution-free, ADF; Browne, 1984), or no distributional assumption (unweighted least squares, ULS; Bollen, 1989). These other methods are less popular, because they do not generally produce inferential statistics (ULS), or require relatively large sample sizes (ADF).

The assumption of a multivariate normal distribution for $\mathbf{y}_i$ in ML estimation has the following implications for the variables in equation (1): η_i is drawn from a normal distribution; the residuals are homoscedastic and normally distributed; and the factor loadings do not differ across individuals. If any of these assumptions is violated, non-normality arises in the observed data. Several tests of multivariate normality have been developed. Mardia (1970) proposed measures of multivariate skewness and kurtosis, and Cox and Small (1978) proposed a test based on non-linearity of the dependencies between observed variables. In addition, some marginal tests are used to consider univariate normality for each variable separately (e.g. Shapiro & Wilks, 1965). From the point of view of the possible sources of non-normality in the one-factor model, these tests are omnibus tests, i.e. the three assumptions mentioned above are tested simultaneously. Thus, if the assumption of multivariate normality is rejected, it is unclear which aspect of the common factor model underlies this violation. That is, it is unclear whether the assumptions on the common factor distribution, the residual variances or the factor loadings are violated. To date, relatively little effort has been devoted to the development of procedures to test these aspects of the assumptions of multivariate normality. The literature that exists on these sources of non-normality is discussed next.

1.1. The common factor distribution

A skewed common factor distribution will result in skewed observed variables. A small number of studies have focused on non-normal common factor distributions. In these models, the distribution of the factor (or latent trait) is approximated by means of a histogram (Muthén & Muthén, 2007; Vermunt, 2004; Vermunt & Hagenaars, 2004) or by means of Gauss–Hermite quadratures, where the weights are estimated freely (Schmitt, Mehta, Aggen, Kubarych, & Neale, 2006). In this way, a non-normal factor distribution is approximated to a degree of accuracy that depends on the number of weights or the resolution of the histogram.

1.2. The residual variances

If the assumption of homoscedastic residuals is violated, the residual variances vary with the common factor. Consequences of hetereoscedasticity in the residuals depend on how the residual variances vary exactly with η. For instance, if η is normally distributed and the residual variances increase with η, the observed variable distributions are positively skewed.

Heteroscedasticity of the residuals has been addressed in a number of models. Meijer and Mooijaart (1996) used GLS based on the first three sample moments to fit models that include heteroscedastic residuals. Lewin-Koh and Amemiya (2003) presented a distribution-free method to model heteroscedasticity, again using the first three sample moments. Bollen (1996) did not rely on sample moments, but used a two stage least squares procedure to fit heteroscedastic models. Finally, Hessen and Dolan (2009) presented a test of heteroscedastic residuals based on MML.

1.3. The factor loadings

An additional manner in which non-normality can arise is through the factor loadings. We consider two possibilities. First, the factor loadings are random, i.e. they differ randomly between individuals. This results in a distribution of the observed variables that departs from normality; plots in Kelderman and Molenaar (2007), suggest that random factor loadings tend to affect the kurtosis of the observed variable distribution. Ansari, Jedidi, and Dube (2002) provided a method based on Markov chain Mote Carlo to fit factor models in which the variance of the factor loadings is estimated. Ansari *et al.* proposed a test of homogeneity in which the variance of the individual factor loadings is tested to be equal to zero using pseudo Bayes factors.

In the present paper, we do not focus on this aspect of the multivariate normal distribution as we accept that the multi-level structure of the data can account for the non-normality of the observed data. Rather, we focus on a second manner in which non-normality arises through the factor loadings. In this case, the factor loadings vary systematically with the common factor. We refer to this as the *level dependency* of the factor loadings. In the case where factor loadings increase (decrease) with η, the observed data will be positively (negatively) skewed. Level dependent factor loadings can also be considered as random factor loadings; however, in contrast with Kelderman and Molenaar (2007) and Ansari *et al.* (2002), the randomness has different consequences for the distribution of the observed data (it affects the skewness instead of the kurtosis), it can be explained by taking η into account, and a possible multi-level structure of the data is unlikely to account for the non-normality.

To our knowledge, level dependency of factor loadings has yet to be considered in the literature. However, we do acknowledge that models with non-linear factor-to-indicator relations give rise to level dependent factor loadings (e.g. Etezadi-Amoli & McDonald, 1983; Jaccard & Wan, 1995; Kenny & Judd, 1984; Klein & Moosbrugger, 2000; Lee & Zhu, 2002; McDonald, 1962; Yalcin & Amemiya, 2001). We elaborate on this below.

This paper focuses on these three effects in the one-factor model: non-normal factor distribution, heteroscedastic residuals, and level dependent factor loadings. Based partly on Hessen and Dolan (2009), we present a unified approach based on MML. This approach is promising, because it enables the test of the distributional assumption underlying factor analysis, i.e. that of a multivariate normal distribution of the observed data; it enables the identification of specific violations of the multivariate normal distribution; and once identified, it allows one to model these violations. The detection of specific violations in the one-factor model could be of theoretical interest from an applied point of view. An important example is ability differentiation, a recurring theme in intelligence research (e.g. Deary *et al.*, 1996). We show below how the present approach provides a general framework for testing this differentiation hypothesis.

The present undertaking is related to work done in the context of univariate regression models for location, scale, and shape (Rigby & Stasinopoulos, 2005).[1] This approach extends the univariate manifest regression model, where the mean of the dependent variable is modelled as a function of the independent variables, by modelling the other parameters of the distribution of the dependent variable as well. In the present paper, we model the higher moments of the distribution of the dependent variable, $\mathbf{y}_i$ as a function of an unobserved independent variable, the common factor. It is therefore related to the approach of Rigby and Stasinopoulos, but is unique in the sense that it concerns an unobserved independent variable.

The organization of this paper is as follows. First, we present the model and show how we estimate the parameters. Next, we present simulation results to demonstrate the viability of the model. Then, we fit the model to IQ data to illustrate the usefulness of the model in investigating ability differentiation. We conclude the paper with a general discussion.

2. The formal model and estimation of the parameters

2.1. Heteroscedastic residuals

Hessen and Dolan (2009) developed a one-factor model that incorporates heteroscedastic residuals. In their approach, they use the information of the whole system of simultaneous equations to improve the efficiency of the estimator rather than using moments (Lewin-Koh & Amemiya, 2003; Meijer & Mooijaart, 1996). They included an explicit test of homoscedasticity subject to the definition of a functional relationship linking the level of η with the variance of the individual residuals (see equation (2) below). Here we generalize this approach. As in Hessen and Dolan (2009), heteroscedasticity is taken into account by modelling the logarithm of the residual variances conditional on η (see also Harvey, 1976). The conditional residual variance of variable j is then given by

$$\log\left(\sigma^2_{\varepsilon j}|\eta\right) = \beta_{j0} + \beta_{j1}\eta + \beta_{j2}\eta^2 + \cdots + \beta_{jr}\eta^r, \tag{2}$$

i.e. the residual variance of variable j conditional on η, $\sigma^2_{\epsilon i}|\eta$, is modelled by an rth-order polynomial function. The parameter β_{j0} accounts for the residual variance in the observed variable that is independent of the factor, and β_{js} for $s = 1, \ldots, r$ accounts for the residual variance that is a function of the factor. Using equation (2), the nature of the heteroscedasticity in the residuals can be modelled to any degree of accuracy, depending on the choice of r. For $r = 1$ a special model arises, which Hessen and Dolan refer to as the minimal heteroscedastic model, i.e.

$$\log\left(\sigma^2_{\varepsilon j}|\eta\right) = \beta_{j0} + \beta_{j1}\eta, \tag{3}$$

where β_{j1} is now interpreted as a heteroscedasticity parameter. This parameter is used to check the assumption of homoscedasticity for variable j by testing

$$H_0 : \beta_{j1} = 0$$

[1] We thank an anonymous reviewer for bringing this paper to our attention.

against

$$H_A : \beta_{j1} \neq 0$$

using a likelihood ratio test or Wald test (see Buse, 1982; Hessen & Dolan, 2009). If H_0 holds, the residual variances are independent of η and the residuals are homoscedastic. This is the test for minimal heteroscedasticity; a more elaborated approach would be to test

$$H_0 : \beta_{j1} = \beta_{j2} = \cdots = \beta_{jr} = 0, \tag{4}$$

for $r > 1$. Note that in both approaches, a likelihood ratio test enables the multivariate test of homoscedasticity for multiple observed variables.

2.2. Non-normality in the factor distribution

Previous research has focused on semi-parametric models, i.e. models in which the distribution of the factor is approximated using a histogram. Two approaches can be distinguished. The first is a latent class approach. In this approach, the bars of the histogram are formulated as latent classes with zero variance within each class. The mean of each class corresponds to the position of the bars on the factor scale, and the size of each class (i.e. the class proportion) corresponds to the height of the bar (Muthén & Muthén, 2007; Vermunt, 2004; Vermunt & Hagenaars, 2004). The second approach uses Gauss–Hermite quadratures. Here, the nodes correspond to the position of the bars, and the weights correspond to the height of the bars. Non-normality is taken into account by estimating the weights, which are normally fixed to certain values (Schmitt *et al.*, 2006). In both approaches, a test of normality involves comparing the length of the bars to those that are expected under the standard normal density function. These tests commonly involve multiple degrees of freedom.

In this paper, we propose a parametric approach that is able to model non-normality in the factor distribution. Tests for normality involve one degree of freedom. Below we note how our approach is related to these semi-parametric models.

Our approach is to adopt the following density for the factor distribution

$$h(\eta|\kappa, \omega, \zeta) = \frac{2}{\omega} \times \Phi\left(\zeta \frac{\eta - \kappa}{\omega}\right) \times \varphi\left(\frac{\eta - \kappa}{\omega}\right), \tag{5}$$

where ω is a scale parameter, κ is a location parameter, ζ is a shape parameter, $\varphi(.)$ refers to the standard normal density function, and $\Phi(.)$ refers to the standard normal distribution function. This distribution is known as the skew-normal distribution, developed by Azzalini (1985, 1986). The parameters κ and ω do not correspond directly to the expected value and variance of the distribution. However, these are calculated using (Azzalini & Capatanio, 1999)

$$E(\eta) = \kappa + \omega\sqrt{\frac{2}{\pi}}\delta \tag{6}$$

and

$$\mathrm{Var}(\eta) = \omega^2\left(1 - \frac{2\delta^2}{\pi}\right), \tag{7}$$

where

$$\delta = \frac{\zeta}{\sqrt{1+\zeta^2}}. \tag{8}$$

An important property of the skew-normal distribution is that it includes the normal distribution as a special case. That is, when $\zeta = 0$, equation (5) reduces to

$$h(\eta|\kappa, \omega, \zeta) = \frac{1}{\omega} \times \varphi\left(\frac{\eta - \kappa}{\omega}\right)$$

which is a normal density function with $E(\eta) = \kappa$ and $\mathrm{Var}(\eta) = \omega^2$.

Equations (5)–(7) form the skew-normal density as we use it here. Although there is some literature concerning the generating process behind the skew-normal distribution (i.e. the density arises from a selection process; see Fernando de Helguero, cited by Azzalini, 2009; Molenaar, 2007), for the present purpose, the generating process is not of interest. Specifically, we use the skew-normal density for reasons of convenience. Most importantly, as the skew-normal density includes the normal distribution as a special case for $\zeta = 0$, a straightforward test for normality is

$$H_0 : \zeta = 0.$$

In addition, statistical properties of the skew-normal distribution are well documented (e.g. Arnold & Beaver, 2002; Arnold, Beaver, Groeneveld, & Meeker, 1993; Azzalini & Capatanio, 1999; Azzalini & Dalla Valle, 1996; Chiogna, 2005; Monti, 2003), the distribution is widely applied (for good examples, see Azzalini, 2005), and in the one-factor model, the distribution is relatively easy to implement. Thus, we use the skew-normal density as a statistical tool to investigate the nature of the common factor distribution. However, our choice is based on pragmatic considerations. Other distributions, possibly substantively motivated, may be considered as well.

2.3. Level dependency of the factor loadings

The level independency of the factor loadings has not been addressed previously, to our knowledge. We propose to take level dependency into account by modelling the factor loadings conditional on η by

$$\lambda_j(\eta) = s(\eta)$$

where $s(.)$ is a suitable function. In the general case, a polynomial function will suffice, i.e.

$$\lambda_j(\eta) = \gamma_{j0} + \gamma_{j1}\eta + \gamma_{j2}\eta^2 + \cdots + \gamma_{jr}\eta^r, \tag{9}$$

which is a polynomial of the rth degree. In this function, γ_{j0} is the baseline factor loading, which is independent of η, and γ_{sj} ($s = 1, \ldots, r$) accounts for the dependency between the factor loadings and the factor. In doing so, the level dependency of the factor loadings can be approximated to a degree of accuracy that depends on the order of the polynomial, r. Note that equation (9) is just one choice for $s(.)$. Other functions could be used instead, as we demonstrate in Section 4.

As in the model for heteroscedasticity, we propose a minimal level dependency model in the polynomial model from equation (9). By setting $r = 1$, we obtain

$$\lambda_j(\eta) = \gamma_{j0} + \gamma_{j1}\eta, \tag{10}$$

where γ_{j1} is interpreted as a level dependency parameter. To test the assumption that the factor loading of observed variable j is level independent involves testing whether

$$H_0 : \gamma_{j1} = 0$$

holds. If H_0 is rejected the factor loading of variable j is level dependent. This test of level dependency could be extended to include higher-order terms (as in equation (4)). In addition, as noted above, multiple observed variables could be tested simultaneously using a likelihood ratio test.

To test the minimal level dependency model from equation (10), existing models could be used as well. By substituting equation (10) in equation (1) we obtain

$$y_{ij} = \upsilon_j + \gamma_{j0}\eta_i + \gamma_{j1}\eta_i^2 + \varepsilon_{ij}, \tag{11}$$

which can be fitted using non-linear factor analyses (e.g. Bauer, 2005a; Klein & Moosbrugger, 2000; Lee & Zhu, 2002; Yalcin & Amemiya, 2001). Given present purposes, we see some advantages of our approach over non-linear factor analyses. First, as we model the factor loadings conditional on η, we do not impose a normality assumption on the unconditional observed data. This normality assumption is a recurrent problem in non-linear factor analyses (for instance, see Klein & Moosbrugger, 2000). Second, using non-linear factor analyses it is not straightforward to fit models like equation (9) where fitting these models in our approach does not pose a problem. Third, in non-linear factor analyses there is limited flexibility with respect to the function between the factor loadings and the factor. For instance, in the present model it is straightforward to implement a logistic function, as we demonstrate in section 4. Fourth, in modelling level dependency in the factor loadings, it may be desirable to include heteroscedastic residuals, which is possible in the present approach.

Due to its equivalence to non-linear factor analysis, the level dependent factor model has important implications for measurement invariance (Meredith, 1993). Bauer (2005b) pointed out that when the true factor-to-indicator relation is quadratic and invariant across groups, testing for measurement invariance in the linear factor model will result in different factor loadings and intercepts across those groups. These differences increase as a function of the common factor mean difference. It is therefore important to test the assumption of linearity and to take into account possible non-linearity when proceeding in the test of measurement invariance. The level dependent factor loadings model presented here is a suitable vehicle for this.

2.4. Marginal maximum likelihood

Using equations (1), (2), (5), and (9), we obtain a common factor model with heteroscedastic residuals, level dependent factor loadings, and a skew-normal factor distribution. We use MML to fit this model (Bock & Aitkin, 1981; see also Hessen & Dolan, 2009; van der Sluis, Dolan, Neale, Boomsma, & Posthuma, 2006).

The log-likelihood function of the model is

$$\log L(\boldsymbol{\tau}|\mathbf{y}) \approx \sum_{i=1}^{N} \log\left(\sum_{q=1}^{Q} W_q^* \times f(\mathbf{y}_i | N_q^*, \boldsymbol{\tau}) \right), \tag{12}$$

where

$$W_q^* = \frac{2}{\sqrt{\pi}} \times W_q \times \Phi(\zeta\sqrt{2}\omega N_q) \tag{13}$$

and

$$N_q^* = \sqrt{2}\omega \times N_q + \kappa. \tag{14}$$

W_q and N_q are transformations of the Q 'weights' and 'nodes' from a Gauss–Hermite guadrature approximation of the integral in the marginal likelihood function, N is the number of subjects in the sample, $\boldsymbol{\tau}$ is the vector of parameters to be estimated (e.g. υ_j, β_{j0}, β_{j1}, γ_{j0}, γ_{j1}, $E(\eta)$, $\mathrm{Var}(\eta)$, and ζ), and $f(.)$ is given by

$$f(\mathbf{y}_i|\eta, \boldsymbol{\tau}) = \frac{1}{\prod_{j=1}^{p} \sqrt{2\pi \exp\left(\sum_{s=0}^{r} \beta_{js}\eta^s\right)}} \exp\left(-\frac{1}{2} \sum_{j=1}^{p} \frac{\left[y_{ij} - \left(\upsilon_j + \sum_{s=0}^{r} \gamma_{js}\eta^{s+1} \right) \right]^2}{\exp\left(\sum_{s=0}^{r} \beta_{js}\eta^s \right)} \right).$$

See the Appendix for a derivation of the log-likelihood function. It is important to note that the log-likelihood in equation (12) is closely connected to a latent class approach (e.g. Vermunt, 2004). W_q^* are weights between 0 and 1 that sum to 1; these therefore correspond to Q class proportions. N_q^* are the nodes that are symmetric and sum to 0; these therefore correspond to the Q class means. The major difference between a latent class approach and our approach is that in the latent class approach, the class proportions and means are estimated, while in our approach these are fixed and transformed according to some function (i.e. equations (13) and (14)). We only estimate the parameters from this transformation, instead of all W_q and N_q. The log-likelihood function can also be formulated in terms of a mixture distribution. Then, we have Q mixtures of $f(\mathbf{y}_i|\eta, \tau)$, with W_q^* as the mixing proportions, and N_q^* as the component means of η. By formulating the model in this way, it is straightforward to fit it to data using the freely available software package Mx (Neale, Boker, Xie, & Maes, 2002), but also in Mplus (Muthén & Muthén, 2007).

2.5. Identification

As it stands, the model is underidentified. We impose two restrictions to identify the model. First, the traditional scale restriction is imposed (e.g. see Bollen, 1989, p. 238). That is, γ_{j0} of one variable or $\mathrm{Var}(\eta)$ is restricted to equal some non-zero constant (commonly 1). Second, a metric restriction is imposed, that is, $E(\eta)$ is constrained to equal some constant (commonly 0). Alternatively, ν_j is restricted to some known constant (commonly 0) for some j.

Given these restrictions, all effects can be estimated separately, but not necessarily simultaneously. We relied on empirical identification (i.e. confidence intervals, the rank of the Hessian, and simulation results) to determine which effects can be combined given realistic sample sizes. It turned out that the parameters from the minimal heteroscedastic model and the parameters from the minimal level dependency model can be estimated simultaneously. That is, β_{j1} and γ_{j1} can be estimated together for all $j = 1, \ldots, p$. Furthermore, the parameters from the minimal heteroscedastic model and the shape parameter can be estimated simultaneously. That is, β_{j1} and ζ can be estimated together for all $j = 1, \ldots, p$. However, it is not possible to estimate the parameters from the minimal level dependency model together with the shape parameter. This results in an ill-conditioned model. See Section 3 for some details on the empirical identification.

An issue related to identification is the curvature of the log-likelihood function. First, it is known that the log-likelihood function has a stationairy point for $\zeta = 0$ (see Azzalini, 1985). Second, it appears that with few data and a large true value for ζ, the log-likelihood is monotome increasing for $\zeta \rightarrow \infty$ (see Chiogna, 2005; Monti, 2003). The first anomaly is overcome by assigning non-zero starting values to ζ, and by refitting the model with different starting values for this parameter to see whether it makes a difference. The second anomaly did not appear to be a major problem, as is demonstrated in the next section. Only in 3 out of 1,800 cases ζ was estimated on its boundary. If such a situation arises in practice, one may conclude that the number of individuals is not sufficient to estimate the skewness in the common factor distribution.

3. Simulation studies

We establish the viability of the model by two small simulation studies. The first simulation study is intended to see whether parameters are recovered adequately under a number of circumstances. The second simulation study is intended to see how the different effects are absorbed in the parameters of the other effects.

3.1. Simulation study 1

3.1.1. Design

We simulated data according to the model defined by equations (1), (3), (5), and (10). Note that we rely on the minimal heteroscedastic model and the minimal level dependency model, as we want to focus on hypothesis testing as opposed to statistical modelling. In this undertaking, we manipulated the number of subjects (3 levels: 300, 400, or 500 subjects); the nature of the effects present (4 levels: all effects separate, i.e. $\gamma_{j1} \neq 0$, $\zeta \neq 0$, and the combinations, i.e. $\beta_{j1} \neq 0$, $\gamma_{j1} \neq 0$, and $\beta_{j1} \neq 0$, $\zeta \neq 0$);[3] and the size of the effects (3 levels: small, medium, large). These manipulations gave rise to 36 conditions ($3 \times 4 \times 3$). Aspects that were not manipulated concerned the number of variables, which was fixed to equal 5; the intercepts (υ_j from equation (1)) which were fixed to equal 1; and the baseline parameters β_{j0} and γ_{j0}, which were fixed to equal 0.5 and 1, respectively, for all $j = 1, \ldots, 5$. Finally, $E(\eta)$ and $\mathrm{Var}(\eta)$ were fixed to equal 0 and 1, respectively. Values of the parameters γ_{j1}, β_{j1}, and ζ were manipulated as described next.

In the case of a skew-normal factor distribution ($\zeta \neq 0$), effect size was defined as the size of the third standardized moment (skewness). The skewness of the skew-normal

[3] Notice that in simulation study 1, we do not investigate heteroscedastic residuals in isolation, as this model is equivalent to the model of Hessen and Dolan (2009). We refer to the latter paper for results limited to these effects.

density (α_1) is calculated according to Azzalini (1985) as

$$\alpha_1 = \frac{4-\pi}{2}\frac{(\delta\sqrt{2/\pi})^3}{(1-2\delta^2/\pi)^{3/2}},$$

where δ is given by equation (8). We choose this coefficient to be either 0.4 (small, i.e. $\zeta = 1.81$), 0.5 (medium, i.e. $\zeta = 2.17$), or 0.6 (large, i.e. $\zeta = 2.62$). In the case of level dependent factor loadings ($\gamma_{j1} \neq 0$), all γ_{j1} from the minimal level dependency model were either 0.1 (small), 0.2 (medium), or 0.3 (large). In the case of heteroscedastic residuals ($\beta_{j1} \neq 0$), all β_{j1} from the minimal heteroscedastic model are either 0.1 (small), 0.2 (medium), or 0.3 (large). To get some idea of the effect size of the manipulations of γ_{j1} and β_{j1}, we show in Figure 1 how the residual variances and the factor loadings vary as a function of η. It appears that in the case of a small effect size, both the factor loadings and the residual variances increase only slightly across η, while in the case of medium and large effect sizes, they increase more.

For each model, we determined the empirical power to detect the effects in the model using the likelihood ratio test (Saris & Satorra, 1993; Satorra & Saris, 1985). In this procedure, the non-centrality parameter of the χ^2 distribution of the likelihood ratio test statistic is estimated under the restricted model (i.e. the model that does not include the effect of interest). For details we refer the reader to Saris and Satorra (1993) and Satorra and Saris (1985).

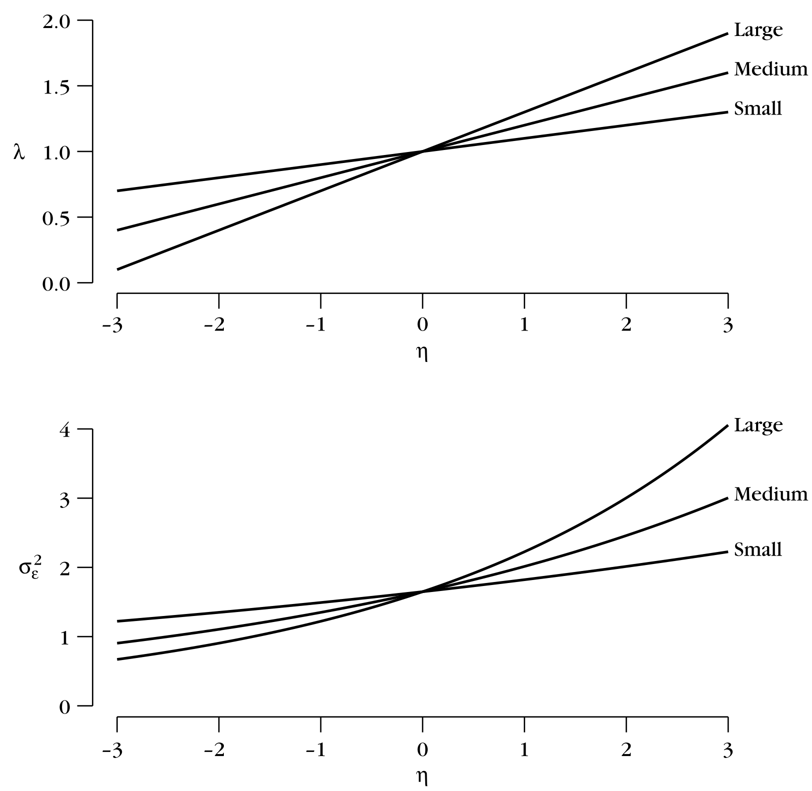

Figure 1. The factor loadings (top) and the residual variances (bottom) as a function of η for the small, medium, and large effect size condition.

For each of the 36 conditions 100 data sets were simulated and the true model was fitted to these data. Parameters were estimated in Mx (Neale *et al.*, 2002) by minimizing -2 times the log-likelihood function of the model (equation (12)). In all subsequent analyses we will use 100 quadrature points (i.e. $Q = 100$). The Mx input files of all models are available from the website of the first author.

3.1.2. Results

In Table 1, the means and standard deviations of the estimates of the manipulated parameters are shown for each condition.[4] Judged by these results, parameter values are generally recovered quite well, and are quite unbiased.

In Table 2, rejection rates of the Shapiro–Wilks test of normality are shown for each condition. For instance, a rejection rate of 32 for variable 1 denotes that within that condition, the null hypothesis of a normal distribution for the scores on variable 1 is rejected in 32 of the 100 data sets given an alpha of .01.

As effect sizes and the number of subjects increase, rejection rates increase, as expected. It is notable that, for small effect sizes and relatively few subjects ($N = 300$), rejection rates are around the Type I error rate (.01), while the results from Table 1 show that the model with $\zeta \neq 0$ detects a significant amount of skewness. That is, the observed data appear to be normally distributed according to the Shapiro–Wilks normality test, while there is a small amount of skewness that is detectable using the model. Table 2 also lists empirical power coefficients for each model under each condition. These power coefficients could be interpreted as the statistical power to reject the null hypothesis of no effects, e.g. for a model with $\gamma_{j1} \neq 0$, the reported power coefficient is interpreted as the power to reject $H_0 : \gamma_{j1} = 0$, for all $j = 1, \ldots, 5$. For models that exist of combined effects, two power coefficients are reported. The first coefficient denotes the power associated with the effect that is listed first (i.e. level dependency of the factor loadings or non-normality of the factor distribution). The second coefficient denotes the power to detect both effects simultaneously, i.e. the power to reject the null hypothesis:

$$H_0 : \beta_{j1} = \zeta = 0, \quad \text{for all } j = 1, \ldots, 5.$$

As seen from Table 2, an interesting result is that as effect sizes and the number of subjects increase, rejection rates increase to close to 100, except for the model with $\zeta \neq 0$. Rejection rates for data generated according to this model are 36 at most. This indicates that the Shapiro–Wilks normality test has little power to detect non-normality due to the common factor distribution. However, as one can see in Table 2, the power of the null hypothesis that $\zeta = 0$ equals 0.99, which is large.

3.2. Simulation study 2

3.2.1. Design

In this study, we simulated 100 data sets according to each of three models: a model with heteroscedastic residuals ($\beta_{j1} \neq 0$); a model with level dependent factor loadings

[4] In two cases, ζ was estimated on its boundary. One case was in the condition with $\zeta \neq 0$, $N = 300$ and a large effect size, and the other case was in the condition with $\beta_{j1} \neq 0$ and $\zeta \neq 0$, $N = 400$ and a medium effect size. Both cases were removed from the results.

Table 1. Parameter estimates for simulation study 1

Effect	N	Model	ζ	γ_{11}	γ_{21}	γ_{31}	γ_{41}	γ_{51}	β_{11}	β_{21}	β_{31}	β_{41}	β_{51}
Small			*1.81*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*
	300	$\gamma_{j1} \neq 0$	–	0.09 (0.07)	0.10 (0.07)	0.09 (0.06)	0.09 (0.08)	0.11 (0.07)	–	–	–	–	–
		$\zeta \neq 0$	1.83 (0.73)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.10 (0.08)	0.11 (0.08)	0.10 (0.07)	0.11 (0.08)	0.12 (0.08)	0.11 (0.12)	0.10 (0.11)	0.12 (0.11)	0.09 (0.13)	0.11 (0.10)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	1.90 (0.71)	–	–	–	–	–	0.06 (0.07)	0.06 (0.08)	0.07 (0.07)	0.07 (0.07)	0.06 (0.07)
	400	$\gamma_{j1} \neq 0$	–	0.09 (0.06)	0.10 (0.06)	0.11 (0.06)	0.10 (0.06)	0.10 (0.06)	–	–	–	–	–
		$\zeta \neq 0$	1.92 (0.57)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.11 (0.08)	0.11 (0.06)	0.09 (0.06)	0.10 (0.07)	0.10 (0.06)	0.10 (0.08)	0.08 (0.11)	0.11 (0.11)	0.12 (0.09)	0.11 (0.10)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	1.90 (0.62)	–	–	–	–	–	0.06 (0.06)	0.07 (0.05)	0.07 (0.06)	0.07 (0.06)	0.07 (0.06)
	500	$\gamma_{j1} \neq 0$	–	0.10 (0.06)	0.10 (0.05)	0.09 (0.06)	0.10 (0.06)	0.09 (0.05)	–	–	–	–	–
		$\zeta \neq 0$	1.75 (0.39)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.11 (0.06)	0.10 (0.06)	0.09 (0.05)	0.10 (0.05)	0.10 (0.06)	0.10 (0.08)	0.10 (0.09)	0.10 (0.09)	0.10 (0.08)	0.12 (0.08)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	1.82 (0.46)	–	–	–	–	–	0.06 (0.06)	0.08 (0.05)	0.07 (0.05)	0.07 (0.05)	0.06 (0.05)
Medium			*2.17*	*0.2*	*0.2*	*0.2*	*0.2*	*0.2*	*0.2*	*0.2*	*0.2*	*0.2*	*0.2*
	300	$\gamma_{j1} \neq 0$	–	0.20 (0.07)	0.20 (0.07)	0.21 (0.07)	0.20 (0.08)	0.20 (0.08)	–	–	–	–	–
		$\zeta \neq 0$	2.19 (0.61)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.20 (0.08)	0.20 (0.08)	0.20 (0.07)	0.20 (0.07)	0.20 (0.08)	0.20 (0.12)	0.21 (0.11)	0.19 (0.11)	0.19 (0.14)	0.21 (0.11)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	2.26 (0.74)	–	–	–	–	–	0.13 (0.07)	0.13 (0.07)	0.13 (0.08)	0.12 (0.07)	0.14 (0.07)
	400	$\gamma_{j1} \neq 0$	–	0.20 (0.07)	0.21 (0.07)	0.21 (0.07)	0.21 (0.07)	0.20 (0.07)	–	–	–	–	–
		$\zeta \neq 0$	2.29 (0.76)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.18 (0.08)	0.21 (0.07)	0.20 (0.06)	0.19 (0.07)	0.20 (0.07)	0.21 (0.11)	0.19 (0.09)	0.19 (0.11)	0.21 (0.10)	0.21 (0.09)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	2.15 (0.51)	–	–	–	–	–	0.14 (0.06)	0.13 (0.05)	0.13 (0.06)	0.13 (0.05)	0.13 (0.05)
	500	$\gamma_{j1} \neq 0$	–	0.20 (0.06)	0.20 (0.06)	0.19 (0.05)	0.20 (0.06)	0.19 (0.06)	–	–	–	–	–
		$\zeta \neq 0$	2.30 (0.58)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.21 (0.06)	0.20 (0.06)	0.21 (0.07)	0.20 (0.07)	0.21 (0.06)	0.20 (0.09)	0.22 (0.08)	0.18 (0.10)	0.21 (0.08)	0.19 (0.10)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	2.21 (0.53)	–	–	–	–	–	0.14 (0.05)	0.13 (0.06)	0.14 (0.05)	0.14 (0.05)	0.13 (0.06)
Large			*2.62*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*
	300	$\gamma_{j1} \neq 0$	–	0.30 (0.10)	0.30 (0.11)	0.30 (0.10)	0.31 (0.10)	0.31 (0.11)	–	–	–	–	–
		$\zeta \neq 0$	2.91 (1.12)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.37 (0.21)	0.37 (0.18)	0.36 (0.20)	0.38 (0.20)	0.39 (0.22)	0.33 (0.14)	0.33 (0.14)	0.34 (0.15)	0.30 (0.12)	0.32 (0.14)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	2.80 (0.91)	–	–	–	–	–	0.20 (0.07)	0.19 (0.07)	0.20 (0.06)	0.19 (0.07)	0.18 (0.07)
	400	$\gamma_{j1} \neq 0$	–	0.31 (0.09)	0.31 (0.08)	0.30 (0.09)	0.30 (0.11)	0.30 (0.08)	–	–	–	–	–
		$\zeta \neq 0$	2.89 (1.02)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.31 (0.12)	0.32 (0.11)	0.31 (0.11)	0.33 (0.11)	0.31 (0.13)	0.30 (0.11)	0.30 (0.09)	0.32 (0.11)	0.28 (0.10)	0.31 (0.09)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	2.58 (0.71)	–	–	–	–	–	0.19 (0.06)	0.20 (0.06)	0.20 (0.06)	0.20 (0.06)	0.20 (0.06)
	500	$\gamma_{j1} \neq 0$	–	0.30 (0.07)	0.30 (0.07)	0.30 (0.07)	0.30 (0.06)	0.30 (0.07)	–	–	–	–	–
		$\zeta \neq 0$	2.78 (0.74)	–	–	–	–	–	–	–	–	–	–
		$\gamma_{j1} \neq 0$ and $\beta_{j1} \neq 0$	–	0.31 (0.09)	0.31 (0.09)	0.30 (0.09)	0.31 (0.09)	0.30 (0.10)	0.29 (0.09)	0.31 (0.08)	0.32 (0.10)	0.30 (0.09)	0.30 (0.09)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	2.81 (0.67)	–	–	–	–	–	0.19 (0.05)	0.19 (0.05)	0.19 (0.05)	0.20 (0.05)	0.20 (0.06)

Note. True parameter values are in italic. Dashes mean that the corresponding parameter is not in the model (i.e. fixed to zero). Standard deviations are in brackets.

Table 2. Rejection rates and empirical power coefficients for simulation study 1

			Rejection rate item number					Power	
Effect	N	Model	1	2	3	4	5	1	1 and 2
Small	300	$\gamma_{j1} \neq 0$	9	5	6	7	12	0.36 (4.91)	–
		$\zeta \neq 0$	9	2	5	9	9	0.60 (4.95)	–
		$\beta_{11} \neq 0$ and $\gamma_{j1} \neq 0$	16	24	19	18	21	0.68 (10.11)	0.77 (12.01)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	22	21	15	20	31	0.65 (5.45)	0.73 (11.87)
	400	$\gamma_{j1} \neq 0$	3	10	10	6	5	0.48 (6.62)	–
		$\zeta \neq 0$	7	8	10	8	8	0.78 (7.41)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	27	25	31	23	34	0.77 (12.14)	0.83 (13.80)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	27	28	30	36	31	0.76 (7.09)	0.86 (15.64)
	500	$\gamma_{j1} \neq 0$	6	8	13	13	12	0.57 (8.11)	–
		$\zeta \neq 0$	10	13	9	10	9	0.77 (7.32)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	39	41	40	32	46	0.83 (13.62)	0.90 (16.24)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	32	45	33	30	30	0.81 (8.13)	0.93 (19.39)
Medium	300	$\gamma_{j1} \neq 0$	46	38	37	36	28	0.96 (20.76)	–
		$\zeta \neq 0$	9	13	12	9	8	0.77 (7.30)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	71	75	76	71	76	0.98 (24.78)	0.99 (28.76)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	54	54	42	49	52	0.80 (7.78)	1.00 (31.64)
	400	$\gamma_{j1} \neq 0$	39	50	47	51	48	0.99 (27.79)	–
		$\zeta \neq 0$	11	19	24	18	12	0.91 (11.09)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	84	95	90	87	91	1.00 (30.99)	1.00 (37.84)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	70	68	59	63	67	0.87 (9.63)	1.00 (41.03)
	500	$\gamma_{j1} \neq 0$	62	55	49	55	52	1.00 (32.62)	–
		$\zeta \neq 0$	25	33	18	26	23	0.97 (14.89)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	95	97	95	93	99	1.00 (38.08)	1.00 (49.54)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	80	76	76	84	75	0.95 (12.98)	1.00 (54.74)
Large	300	$\gamma_{j1} \neq 0$	75	68	63	66	66	1.00 (41.36)	–
		$\zeta \neq 0$	24	16	15	16	19	0.94 (12.31)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	99	97	100	100	100	1.00 (51.23)	1.00 (62.92)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	85	80	87	80	74	0.92 (11.43)	1.00 (62.87)
	400	$\gamma_{j1} \neq 0$	82	87	79	81	89	1.00 (57.44)	–
		$\zeta \neq 0$	24	27	23	25	24	0.98 (16.31)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	100	100	100	100	100	1.00 (63.53)	1.00 (80.35)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	85	94	94	96	92	0.96 (13.99)	1.00 (86.89)
	500	$\gamma_{j1} \neq 0$	93	91	94	95	95	1.00 (73.42)	–
		$\zeta \neq 0$	25	35	33	30	36	0.99 (20.14)	–
		$\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$	100	100	100	100	100	1.00 (79.92)	1.00 (94.91)
		$\zeta \neq 0$ and $\beta_{j1} \neq 0$	95	98	98	98	98	1.00 (20.64)	1.00 (111.40)

Note. 'Power 1' denotes the power to detect the effect that is listed first (i.e. level dependency or non-normal factor distribution). 'Power 1 and 2' denotes the power to detect both effects simultaneously. Empirical non-centrality parameters are in brackets.

($\gamma_{j1} \neq 0$); and a model with a non-normal factor distribution ($\zeta \neq 0$). In these models, the number of variables was 5, the number of subjects equalled 400, and all other parameters were set to equal those in simulation study 1. Finally, all effect sizes were of medium size (see simulation study 1). To these data, first, two models are fitted: a model with $\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$; and a model with both $\zeta \neq 0$ and $\beta_{j1} \neq 0$. Next, one of the two effects was dropped to gauge the effects on the remaining parameters.

3.2.2. Results
In Table 3, means and standard deviations of the estimated parameters are shown.[5] It appears that when the data are generated according the model with $\zeta \neq 0$, this effect is detected by the shape parameter, ζ, and not by the heteroscedasticity parameters, β_{j1}. However, when fitting a model with $\gamma_{j1} \neq 0$ to the same data, the effect is evident in the γ_{j1} parameters, i.e. the factor loadings show level dependencies. The same applies in the opposite case: when the data are generated according to a model with $\gamma_{j1} \neq 0$, the effect is evident in ζ, but not in β_{j1}. These results show why a model with both $\gamma_{j1} \neq 0$ and $\zeta \neq 0$ is ill conditioned; the effects are statistically not separable in the present approach.

4. Application
4.1. Background
A well-established phenomenon in intelligence research is that all subtest scores of an IQ test (e.g. the Wechsler Adult Intelligence Scale; WAIS) are positively correlated, although they concern distinct cognitive abilities (e.g. spatial ability, verbal ability). This phenomenon is explained by positing a higher-order factor that is assumed to underlie all subtest scores. This factor is the general intelligence factor, or *g* (Jensen, 1998). Ability differentiation refers to the claim that the influence of *g* is not uniform across its range, causing inter-subtest correlations to be larger at the lower end of *g*, and smaller towards the higher end of *g* (Spearman, 1927). The ability differentiation hypothesis has enjoyed a good deal of attention in recent decades, resulting in many papers (e.g. Abad, Colom, Juan-Espinosa, & Garcia, 2003; Carlstedt, 2001; Deary *et al.*, 1996; Detterman & Daniel, 1989; Facon, 2004, 2008; Fogarty & Stankov, 1995; Hartmann & Teasdale, 2004; Jensen, 2003; Reynolds & Keith, 2007). All of these papers are based on the creation of subgroups that differ on average in their position on the *g* scale. Most authors used observed subtest scores to create these groups. Under ideal circumstances this method will be adequate. However, forming groups on basis of observed scores (the dependent variables) can distort the factor structure as established in the total population (Muthén, 1989). It would therefore be more adequate to create subgroups on basis of *g* (the independent variable), but this is hard to do as *g* is an unobserved variable. Reynolds and Keith (2007) used Anderson–Rubin factor scores to estimate *g* scores and median-split these to obtain two groups. Anderson–Rubin factor scores are known to be unbiased and structure preserving (Jöreskog, Sörbom, du Toit, & du Toit, 1999; Saris, de Pijper, & Mulder, 1978). This method may be viable, but we argue that the creation of subgroups is not necessary given the present approach. We see three – not mutually exclusive – possibilities, in which ability differentiation could arise. That is, there are three possible effects in the common factor model that make subtest correlations lower for higher values on the common factor. The first possibility is that ability differentiation is caused by a negatively skewed distribution of *g*, with more *g* variance at the lower end of *g*. Thus, we expect $\zeta < 0$. Second, as suggested by Hessen and Dolan (2009), ability differentiation could be due to the heteroscedastic residuals, with larger residual variances at the higher levels of *g*, i.e. $\beta_{j1} < 0$ for all $j = 1, \ldots, p$. Finally, ability differentiation could be caused by level dependency in the factor loadings, with smaller

[5] In one case, ζ was estimated on its boundary. It occurred in the case where the data were simulated with $\gamma_{j1} = 0.2$, while a model with $\zeta \neq 0$ and $\beta_{j1} \neq 0$ was fitted. This case was removed from the results.

Table 3. Parameter estimates for simulation study 2

Data	Parameter estimates										
	ζ	γ_{11}	γ_{21}	γ_{31}	γ_{41}	γ_{51}	β_{11}	β_{21}	β_{31}	β_{41}	β_{51}
$\zeta = 2.17$	2.20 (0.57)	–	–	–	–	–	−0.01 (0.06)	0.00 (0.06)	0.01 (0.06)	−0.01 (0.08)	0.01 (0.07)
	–	0.12 (0.07)	0.12 (0.06)	0.12 (0.06)	0.13 (0.07)	0.12 (0.06)	0.00 (0.10)	0.00 (0.09)	0.02 (0.10)	−0.02 (0.11)	0.02 (0.11)
		–	–	–	–	–	0.01 (0.05)	0.01 (0.04)	0.02 (0.04)	0.00 (0.04)	0.01 (0.04)
$\gamma_{j1} = 0.2$	3.65 (1.17)	–	–	–	–	–	−0.01 (0.06)	0.00 (0.06)	0.01 (0.06)	−0.02 (0.05)	−0.01 (0.06)
	–	0.19 (0.07)	0.21 (0.07)	0.19 (0.08)	0.19 (0.07)	0.19 (0.06)	0.00 (0.10)	−0.01 (0.10)	0.02 (0.10)	−0.02 (0.09)	−0.01 (0.10)
		–	–	–	–	–	0.01 (0.04)	0.01 (0.04)	0.02 (0.05)	0.00 (0.04)	0.00 (0.04)
$\beta_{j1} = 0.2$	0.20 (0.88)	–	–	–	–	–	0.13 (0.05)	0.13 (0.06)	0.14 (0.06)	0.14 (0.06)	0.15 (0.06)
	–	0.01 (0.07)	0.01 (0.06)	0.00 (0.07)	0.00 (0.07)	0.00 (0.06)	0.19 (0.08)	0.20 (0.10)	0.21 (0.10)	0.22 (0.08)	0.22 (0.09)
	–	0.01 (0.05)	0.02 (0.03)	0.01 (0.04)	0.01 (0.05)	0.02 (0.04)	–	–	–	–	–

Note. Dashes mean that the corresponding parameter is not in the model (i.e. fixed to zero). Standard deviations are in brackets.

factor loadings at the higher levels of g, i.e. $\gamma_{j1} < 0$ for all $j = 1, \ldots, p$. Note that – to support ability differentiation – all effects should all be found in the hypothesized direction. If the effects vary in their directions, ability differentiation is unlikely to be the cause.

In this application, we specify different functions between λ and η in the minimal level dependency model, to show that we are not restricted to a linear function. First, we considered the common linear function from equation (10) as a reference model. The disadvantage of this function is that for large values of η relative to γ_{j0}, the $\lambda_j(\eta)$ assume negative values. In modelling intelligence, we want to avoid this so as to retain the positive manifold at all levels of η. To preserve the positive manifold, we specify a function between λ and η that is uniformely positive, i.e.

$$\lambda_j(\eta) = \exp(\gamma_{j0} + \gamma_{j1}\eta_i).$$

This exponential function has a horizontal asymptote of 0 and therefore precludes negative factor loadings. A third option would be to specify a logistic function, i.e.

$$\lambda_j(\eta) = \gamma_{j0} + \gamma_{j1}[1 + \exp(-\eta_i)]^{-1}.$$

This function has the advantage that the factor loadings are bounded. The lower bound is equal to γ_{j0} and the upper bound is equal to $\gamma_{j0} + \gamma_{j1}$. A disadvantage is that, for this function, a linear transformation of η is not allowed, as the function will not maintain its present form. Results from the model are thus difficult to interpret as they depend on the scale of η. This problem does not occur in the other functions specified throughout this paper (e.g. equations (2) and (9)).

4.2. Results

We analysed data from the National Longitudinal Survey of Youth 1997.[6] From the 12 subtests of the Armed Services Vocational Aptitude Battery (ASVAB), we selected five subtests: Arithmetic Reasoning (AR), Paragraph Comprehension (PC), Auto Information (AI), Mathematics Knowledge (MK), Assembling Objects (AO). As the ASVAB is administered adaptively, the scores on these subtests are estimated by the adaptive procedure. The means, standard deviations, and ranges of the subtest sores are depicted in Tables 4 and 5.

The total sample size consisted of 7,127 subjects. For illustrative purposes, we randomly selected 500 subjects. Of this sample, 8 subjects had one or more missing values on the 5 subtests. However, this poses no problem as Mx can handle missing data.

We first investigated whether multivariate normality was tenable. The test of Mardia (1970) showed that there was a significant amount of skewness in the data ($M1 = 103.62$, $p = 1 \times 10^{-8}$), but no excessive kurtosis ($M2 = 1.89$, $p = .06$) as compared to a multivariate normal distribution. The Shapiro–Wilks test showed that for all subtests but MK univariate normality was rejected at a .01 significance level (AR, $W = 0.99, p < .001$; PC, $W = 0.98, p < .001$; AI, $W = 0.99, p = .003$; MK, $W = 0.99$, $p = .028$; AO, $W = 0.98$, $p < .001$). We first fitted the standard one-factor model, as

[6] Retrieved from http://www.nlsinfo.org/web-investigator/ on March 30, 2008.

Table 4. Means, standard deviations, and ranges of the variables in the application.

	AR	PC	AI	MK	AO
Mean	−0.64	−0.45	−1.41	−0.22	−0.63
SD	1.21	1.22	0.75	1.32	1.19
Min	−3.91	−2.99	−3.20	−3.54	−2.93
Max	2.45	2.26	1.44	2.89	2.37

Table 5. Correlations among the variables in the application.

	AR	PC	AI	MK
PC	.73			
AI	.53	.49		
MK	.80	.75	.45	
AO	.60	.58	.38	.59

a baseline model. The model was identified by setting γ_{10} of MK to equal 1,[7] and $E(\eta)$ to equal 0. The fit of the model was acceptable ($-2\log \mathrm{L} = 6{,}188.91$, $df = 2{,}467$, RMSEA = 0.06).

We started by testing heteroscedasticity of the residuals in isolation (i.e. a model with $\beta_{j1} \neq 0$). Parameter estimates and 99% confidence intervals (based on the profile likelihood; Neale & Miller, 1997; Venzon & Moolgavkar, 1988) are depicted in Table 6. As appears from these results, for the subtests AR and AO homoscedasticity was not tenable (i.e. β_{11} and β_{51} differed significantly from zero). Next, together with the heteroscedasticity in the residuals, we introduced level dependencies in the factor loadings according to the three functions (linear, exponential, and logistic).

Table 6 contains the results of these analyses. In this case, the different functions for $\lambda_j(\eta)$ did not result in different conclusions about level dependency in the data. That is, none of the level dependency parameters, γ_{j1}, were significant, as judged by the 99% confidence intervals. As in the previous analysis, homoscedasticity was rejected for subtests AR and AO as indicated by the significant heteroscedaticity parameter estimates, β_{j1}, for these subtests. However, the heteroscedasticity cannot be reasonably interpreted in terms of ability differentiation, as AR has a negative estimate of β_{j1}, indicating that the residual variance decreases over the range of g (or η), while AO has a positive estimate β_{j1}, indicating that the residual variance of this subtest increases over g (or η).

The next step was to introduce non-normality in the factor distribution to see if there is an effect in the data that is too small to be picked up by the level dependency parameters, but that could be detected by the skew-normal factor distribution. In this model all level dependency parameters were fixed to 0, as they were shown to be non-significant.[8] To avoid possible local minima, we used different starting values for the

[7] In case of the exponential model for the factor loadings, γ_{10} was fixed to log(1) to facilitate interpretation across models.
[8] Note that we could have fixed β_{j1} of subtests PC, AI, and MK to equal zero as well. However, as indicated by an anonymous reviewer, it could be interesting to see if there are differences compared to the previous results.

Table 6. Parameter estimates and likelihood ratio test statistics for the application.

Model	LRT*	*df*	Var(η)	ζ		AR	PC	AI	MK	AO
Baseline			1.17	–	γ_{j1}	–	–	–	–	–
					β_{j1}	–	–	–	–	–
$\beta_{j1} \neq 0$	35.16	5	1.16		γ_{j1}	–	–	–	–	–
					β_{j1}	−0.43	0.00	0.06	−0.16	0.23
						(−0.71, −0.17)	(−0.23, 0.24)	(−0.11, 0.25)	(−0.39, 0.06)	(0.03, 0.44)
$\gamma_{j1} \neq 0$ and	39.54	10	1.17	–	γ_{j1}	−0.02	−0.02	−0.03	0.00	0.04
$\beta_{j1} \neq 0$, lin						(−0.10, 0.06)	(−0.11, 0.07)	(−0.09, 0.04)	(−0.09, 0.10)	(−0.05, 0.13)
					β_{j1}	−0.35	−0.02	0.08	−0.19	0.22
						(−0.74, −0.02)	(−0.25, 0.22)	(−0.10, 0.27)	(−0.46, 0.05)	(0.01, 0.42)
$\gamma_{j1} \neq 0$ and	39.51	10	1.17	–	γ_{j1}	−0.03	−0.02	−0.07	0.00	0.05
$\beta_{j1} \neq 0$, exp						(−0.11, 0.06)	(−0.10, 0.07)	(−0.25, 0.09)	(−0.07, 0.08)	(−0.07, 0.17)
					β_{j1}	−0.35	−0.02	0.08	−0.19	0.22
						(−0.73, −0.02)	(−0.25, 0.22)	(−0.10, 0.28)	(−0.46, 0.05)	(0.01, 0.42)
$\gamma_{j1} \neq 0$ and	23.85	10	1.43	–	γ_{j1}	−0.19	−0.11	−0.12	0.01	0.17
$\beta_{j1} \neq 0$, log						(−0.45, 0.29)	(−0.48, 0.32)	(−0.42, 0.19)	(−0.37, 0.49)	(−0.25, 0.68)
					β_{j1}	−0.28	−0.02	0.07	−0.18	0.19
						(−0.71, 0.01)	(−0.23, 0.21)	(−0.09, 0.25)	(−0.42, 0.04)	(0.00, 0.40)
$\zeta \neq 0$ and	34.45	6	1.16	−0.71	γ_{j1}	–	–	–	–	–
$\beta_{j1} \neq 0$				(−2.32, 1.30)						
					β_{j1}	0.42	0.00	0.07	0.16	0.23
						(−0.72, −0.16)	(−0.23, 0.23)	(−0.11, 0.25)	(−0.39, 0.06)	(0.03, 0.44)

Note. 99% Confidence bounds are in brackets. *The reported likelihood ratio test statistic for each model is the difference in $-2\log L$ between that model and the baseline model.

shape parameter, ζ. In the model, ζ was estimated to be -0.71 which is not significant according to its confidence interval; see Table 6. The overall conclusion is therefore that there is no level dependency in the factor loadings or skewness in the factor distribution. In addition, two subtests showed heteroscedastic residuals, but the direction of these effects were not in line with the hypothesized differentiation effects.

5. Discussion

We presented a unified model to test the distributional assumptions underlying ML estimation in the one-factor model. In addition to testing, the model is useful in modelling any distributional violation that is detected. A particular feature of the model is that multiple effects could be combined in a single model, and tested and/or modelled simultaneously.

Testing and modelling violations of distributional assumptions in any statistical model may be of theoretical interest. To illustrate this, we discussed how an empirical phenomenon, ability differentiation, implies specific violations in the common factor model. Once identified, the presence of the phenomenon was easily tested in a data set using the present model.

In addition to the practical utility of the model, we investigated the performance of the model under a number of circumstances in two simulation studies. In the first simulation study, we showed that given realistic sample sizes, parameter recovery and the power to detect the effects in the data were acceptable. It turned out that when a skew-normal factor distribution underlay the data, this effect was not detected in the majority of the cases using a Shapiro–Wilks normality test on the observed data. The effect was detected with the present model. It is not our purpose to claim that the Shapiro–Wilks test is inferior to the tests proposed in this paper; we wish only to stress that by exploiting the given covariance structure (i.e. the one-factor model), it is possible to detect relatively subtle and specific departures from normality that are more difficult to detect by using marginal tests.

In the second simulation study, we showed that level dependent factor loadings and a skew-normal factor distribution cannot be estimated simultaneously. Nonetheless, we think that both are valuable in their own right. If a skew-normal factor distribution underlies the data, this will cause level dependency in the factor loadings in the same direction for all variables. From this finding, one could infer that a skew-normal factor distribution could be present. However, when only some variables show level dependency, or the variables all show level dependency in opposing directions, then one can be quite sure that the factor distribution does not underlie these effects.

In the application we showed the flexibility of the model in terms of the function that is imposed between the factor loadings and the common factor. This flexibility applies equally to the relation between the residual variances and the common factor, as long as a function is specified that is uniformly positive.

Some limitations associated with the presented work should be noted. First, the model as it stands concerns a one-factor model only, while many psychological phenomena are multidimensional. The model may be generalized to multiple factors by using multivariate Gauss–Hermite quadratures. A well-known problem with these quadratures is, however, that when the number of dimensions exceeds 5, the procedure is practically infeasible (Wood *et al.*, 2002). An alternative is to adopt a Bayesian approach to model fitting. This is straightforward, as the likelihood function of the model presented in this paper could easily be implemented in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000).

A second limitation is that the model is developed in the framework of factor analysis for continuous manifest variables. Many psychometric instruments include dichotomous or polytomous items. A generalization of the present model to discrete manifest data is therefore needed. This could, for instance, be a two-parameter logistic model with level dependent discrimination parameters. This possibility has yet to be pursued.

In this paper, we limited the simulations and application to only five variables. Consequently, the question arises how the model performs with more variables. From a numerical point of few adding variables merely increases computation time, but it should pose no additional problems. From a goodness-of-fit point of view, adding variables results in a model that is less likely to fit the data as a single factor is less likely to underlie all covariances among the variables. Specific pairs of variables may show covariations that are not explained by the common factor. These residual covariances can be incorporated in the present model by freeing the corresponding parameters within the skew common factor or level dependent factor loadings model. In case of the heteroscedastic residual model, it is also possible but less straightforward. As the residual variances depend on the common factor, the possibility exists that the residual covariances vary across the common factor as well. In the Mx syntax on the website of the first author, it is straightforward to incorporate this possibility. However, one has to carefully consider the function that is specified between the residual covariances and the common factor, as the residual covariance matrix should be positive definite over this function.

Finally, a comment is in order concerning the factor distribution. Several authors (Bock & Aitkin, 1981; Muthén, 2008; R. J. Mislevy, cited in Muthén, 2008) have noted that evidence for a non-normal factor distribution is generally weak. However, the present investigation shows that given sufficiently large samples, it is possible to estimate the amount of skewness in the factor distribution. In the presented simulation studies, a negligible fraction of the cases failed due to boundary estimates of the shape parameter. We conclude that a sample size of $N = 300$ is sufficient to estimate the degree of skewness in the factor distribution. Further investigation with only 200 subjects showed that the number of boundary estimates increased a little (up to 10%) but still, the majority of the cases converged at reasonable values. Note that these sample size indications depend on the circumstances we simulated (i.e. the effect sizes and number of variables that were chosen), but we think that these are realistic for scientists who are modelling psychological phenomena.

The present model may serve the purely statistical objective of investigating violations of the assumptions in linear factor analysis. As noted above, such violations can give rise to subtle departures from normality, which may be hard to detect in marginal normality tests (e.g. Wilks–Shapiro). We also identified one substantive area of application, namely ability differentiation testing. Specifically, using the present model, we can formulate and test competing hypotheses concerning the sources, if any, of ability differentiation.

Acknowledgements

The research by Dylan Molenaar was made possible by a grant from The Netherlands Organization for Scientific Research. We thank two anonymous reviewers and the associate editor, Andy Field, for their constructive comments on previous drafts of this paper. All Mx input files are available from the website of the first author, www.dylanmolenaar.nl.

References

Abad, F. J., Colom, R., Juan-Espinosa, M., & Garcia, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence*, *31*, 157–166.

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor models: A Bayesian approach. *Psychometrika*, *67*, 49–78.

Arnold, B. C., & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, *11*, 7–54.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., & Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, *58*, 471–488.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, *46*, 199–208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, *32*, 159–188.

Azzalini, A. (2009). A pioneer. In the skew-normal probability distribution. Retrieved from http://azzalini.stat.unipd.it/SN/.

Azzalini, A., & Capatanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, *Series B*, *61*, 579–602.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726.

Bauer, D. J. (2005a). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *12*, 513–535.

Bauer, D. J. (2005b). The role of nonlinear factor-to-indicator relationships in tests of measurement equivalence. *Psychological Methods*, *10*, 305–316.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1996). A limited-information estimator for LISREL models with or without heteroscedastic errors. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 227–241). Mahwah, NJ: Erlbaum.

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

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *American Statistician*, *36*, 153–157.

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, *36*, 589–609.

Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods & Applications*, *14*, 331–334.

Cox, D. R., & Small, N. T. H. (1978). Testing multivariate normality. *Biometrika*, *65*, 263–272.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*, 105–132.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for the low IQ groups. *Intelligence*, *13*, 349–359.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417–440.

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*, 21–50.

Etezadi-Amoli, J., & McDonald, R. P. (1983). A second generation nonlinear factor analysis. *Psychometrika*, *48*, 315–342.

Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence*, *32*, 391–401.

Facon, B. (2008). How does the strength of the relationships between cognitive abilities evolve over the life span for low-IQ vs high-IQ adults? *Intelligence*, *36*, 339–349.

Fogarty, G. J., & Stankov, L. (1995). Challenging the 'law of diminishing returns'. *Intelligence*, *21*, 157–174.

Hartmann, P., & Teasdale, T. W. (2004). A test of Spearman's 'law of diminishing returns' in two large samples of Danish military draftees. *Intelligence*, *32*, 499–508.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, *44*, 461–465.

Hessen, D. J., & Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood estimation. *British Journal of Mathematical and Statistical Psychology*, *62*, 57–77. doi:10.1348/000711007X248884

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348–357.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence*, *31*, 95–105.

Johnson, W., & Bouchard, T. J. (2004). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393–416.

Jöreskog, K. G., Sörborm, D., du Toit, S., & du Toit, M. (1999). *LISREL 8: New statistical features*. Chicago: Scientific Software.

Kelderman, H., & Molenaar, P. C. M. (2007). The effect of individual differences in factor loadings on the standard factor model. *Multivariate Behaviour Research*, *42*, 435–456.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Lawley, D. N. (1943). The application of the maximum likelihood method to factor analysis. *British Journal of Psychology*, *33*, 172–175.

Lee, S.-Y., & Zhu, H.-T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67*, 189–210.

Lewin-Koh, S., & Amemiya, Y. (2003). Heteroscedastic factor analysis. *Biometrika*, *90*, 85–97.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, *27*, 392–415.

Meijer, E., & Mooijaart, A. (1996). Factor analysis with heteroscedastic errors. *British Journal of Mathematical and Statistical Psychology*, *49*, 189–202.

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, *29*, 223–236.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Molenaar, D. (2007). *Measurement and research department report, 3*. Arnhem: Cito.

Monti, A. C. (2003). A note on the estimation of the skew normal and the skew exponential power distributions. *Metron*, *LXI*, 205–219.

Muthén, B. O. (1989). Dichotomous factor analysis of symptom data. *Sociological Methods & Research*, *18*, 19–65.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2002). *Mx: Statistical modeling* (6th ed.). VCU, Richmond, VA: Author.

Neale, M. C., & Miller, M. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, *27*, 113–120.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence*, *35*, 267–281.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, *54*, 507–554.

Saris, W. E., de Pijper, M., & Mulder, J. (1978). Optimal procedures for estimation of factor scores. *Sociological Methods & Research*, *7*, 85–106.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Newbury Park, CA: Sage.

Satorra, A., & Saris, W. E. (1985). The power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.

Schmitt, J. E., Mehta, P. D., Aggen, S. H., Kubarych, T. S., & Neale, M. C. (2006). Semi-nonparametric methods for detecting latent non-normality: A fusion of latent trait and ordered latent class modeling. *Multivariate Behaviour Research*, *41*, 427–443.

Shapiro, S. S., & Wilks, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Stroud, A. H., & Secrest, D. (1966). *Gaussion quadrature formulas*. Englewodd Cliff, NJ: Prentice-Hall.

van der Sluis, S., Dolan, C. V., Neale, M. C., Boomsma, D. I., & Posthuma, D. (2006). Detecting genotype-environment interaction in monozygotic twin data: Comparing the Jinks and Fulker test and a new test based on marginal maximum likelihood estimation. *Twin Research and Human Genetics*, *9*, 377–392.

Venzon, D. J., & Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics*, *37*, 87–92.

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, *58*, 220–233.

Vermunt, J. K., & Hagenaars, J. A. (2004). Ordinal longitudinal data analysis. In R. C. Hauspie, N. Cameron, & L. Molinari (Eds.), *Methods in human growth research* (pp. 374–393). Cambridge: Cambridge University Press.

Wood, R., Wilson, D. T., Gibbons, R. D., Schilling, S. G., Muraki, E., & Bock, R. D. (2002). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago: Scientific Software International.

Yalcin, I., & Amemiya, Y. (2001). Non linear factor analysis as a statistical method. *Statistical Science*, *16*, 275–294.

Received 27 November 2008; revised version received 30 March 2009

Appendix: Derivation of the log-likelihood function of the model

The model is given by equations (1), (2), (5), and (9). To estimate the parameters from this model, we use MML (Bock & Aitkin, 1981; for applications within the one-factor model, see Hessen & Dolan, 2009; and van der Sluis *et al.*, 2006). MML is based on the multivariate normal distribution conditional on η,

$$f(\mathbf{y}_i, |\eta, \boldsymbol{\tau}) = \frac{1}{\prod_{j=1}^{p} \sqrt{2\pi\sigma_{\varepsilon j}^2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{p} \frac{[y_{ij} - (\upsilon_j + \lambda_j \eta)]^2}{\sigma_{\varepsilon j}^2}\right) \tag{A1}$$

In Equation (A1) all parameters are conditional on η, and $\boldsymbol{\tau}$ is the vector of the parameters to be estimated. Now we introduce the level dependent factor loadings and heteroscedastic residuals by substituting $\sigma_{\varepsilon j}^2|\eta$ and $\lambda_j(\eta)$ from respectively equations (2) and (9), resulting in

$$f(\mathbf{y}_i, |\eta, \boldsymbol{\tau}) = \frac{1}{\prod_{j=1}^{p} \sqrt{2\pi \exp\left(\sum_{s=0}^{r} \beta_{js}\eta^s\right)}} \exp\left(-\frac{1}{2}\sum_{j=1}^{p} \frac{\left[y_{ij} - \left(\upsilon_j + \sum_{s=0}^{r} \gamma_{js}\eta^{s+1}\right)\right]^2}{\exp\left(\sum_{s=0}^{r} \beta_{js}\eta^s\right)}\right) \tag{A2}$$

As η is a nuisance parameter, it is integrated out. Hence we obtain the multivariate marginal density of the observed variables for subject i,

$$k(\mathbf{y}_i|\boldsymbol{\tau}) = \int_{-\infty}^{\infty} f(\mathbf{y}_i|\eta, \boldsymbol{\tau}) \times g(\eta)\mathrm{d}\eta,$$

where $g(\eta)$ is the density function of η and $f(.)$ is given by equation (A2). Commonly, a standard normal density is chosen for $g(\eta)$. We will use the skew-normal density given in equation (5). We therefore obtain

$$k(\mathbf{y}_i|\boldsymbol{\tau}) = \frac{2}{\omega} \times \int_{-\infty}^{\infty} f(\mathbf{y}_i|\eta, \boldsymbol{\tau}) \times \Phi\left(\zeta\frac{\eta - \kappa}{\omega}\right) \times \varphi\left(\frac{\eta - \kappa}{\omega}\right)\mathrm{d}\eta. \tag{A3}$$

Recall that $\varphi(.)$ and $\Phi(.)$ refer to the standard normal density and distribution function, respectively.

We approximate the integral from equation (A3) by using Gauss–Hermite quadratures. The general form of the Gauss–Hermite quadrature approximation is

$$\int_{-\infty}^{\infty} k(x) \times e^{-x^2}\mathrm{d}x \approx \sum_{q=1}^{Q} W_q \times k(N_q), \tag{A4}$$

where, $k(.)$ denotes an arbitrary function, and W_q and N_q are the 'weights' and 'nodes' as found in standard tables (e.g. Stroud & Secrest, 1966). At present, the likelihood function is not in the form of equation (A4). We therefore write it as

$$k(\mathbf{y}_i|\boldsymbol{\tau}) = \frac{2}{\sqrt{2\pi}\omega} \times \int_{-\infty}^{\infty} f(\mathbf{y}_i|\eta, \boldsymbol{\tau}) \times \Phi\left(\zeta\frac{\eta - \kappa}{\omega}\right) \times \exp\left[-\frac{1}{2}\left(\frac{\eta - \kappa}{\omega}\right)^2\right]\mathrm{d}\eta.$$

We transform η using the transformation

$$\eta = \sqrt{2}\omega\eta^* + \kappa,$$

with

$$d\eta = \sqrt{2}\omega(d\eta^*).$$

This results in

$$k(\mathbf{y}_i|\boldsymbol{\tau}) = \frac{2}{\sqrt{\pi}} \times \int_{-\infty}^{\infty} f(\mathbf{y}_i|\eta^*, \boldsymbol{\tau}) \times \Phi(\zeta\sqrt{2}\eta^*) \times e^{-\eta^{*2}} d\eta^*.$$

This expression could be approximated with Gauss–Hermite quadratures, resulting in

$$k(\mathbf{y}_i|\boldsymbol{\tau}) \approx \frac{2}{\sqrt{\pi}} \times \sum_{q=1}^{Q} W_q \times f(\mathbf{y}_i|N_q, \boldsymbol{\tau}) \times \Phi(\zeta\sqrt{2}N_q).$$

By specifying

$$N_q^* = \sqrt{2}\omega \times N_q + \kappa$$

and

$$W_q^* = \frac{2}{\sqrt{\pi}} \times W_q \times \Phi(\zeta\sqrt{2}\omega N_q),$$

the expression reduces to

$$k(\mathbf{y}_i|\boldsymbol{\tau}) \approx \sum_{q=1}^{Q} W_q^* \times f(\mathbf{y}_i|N_q^*, \boldsymbol{\tau}), \tag{A5}$$

which is in the form of equation (A4).

Using equation (A5), the marginal likelihood function for a given sample of N subjects is given by

$$L(\boldsymbol{\tau}|\mathbf{y}) \approx \prod_{i=1}^{N} \sum_{q=1}^{Q} W_q^* \times f(\mathbf{y}_i|N_q^*, \boldsymbol{\tau}). \tag{A6}$$

A final step is to take the logarithm of equation (A6) to obtain the marginal log-likelihood function,

$$\log L(\boldsymbol{\tau}|\mathbf{y}) \approx \sum_{i=1}^{N} \log\left(\sum_{q=1}^{Q} W_q^* \times f(\mathbf{y}_i|N_q^*, \boldsymbol{\tau}) \right).$$