Testing distributional assumptions in psychometric measurement models with substantive applications in psychology

Molenaar, D.

Publication date
2012

2

Testing and Modeling Non-Normality within the Linear Factor Model

Maximum Likelihood estimation in the linear 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, i.e., 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 Chapter, a general approach is presented based on Marginal Maximum Likelihood to enable explicit tests of these assumptions in the one factor model. In addition, the approach is suitable to account for the detected violations, to enable statistical modeling 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.

2.1 Introduction

Over the past hundred 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., Johnson & Bouchard, 2004; Dolan, 2000) 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

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

* This chapter is published as: Molenaar, D., Dolan, C.V., & Verhelst, N.D. (2010). Testing and modeling non-normality within the one factor model. *British Journal of Mathematical and Statistical Psychology*, *63*, 293-317.

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 j-th intercept, λ_j represents the j-th 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) and Maximum Likelihood (ML; Lawley, 1943; 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 of $\mathbf{y_i}$ in ML estimation implies the following for the variables in Equation 1: 1) η_i is drawn from a normal distribution; 2) the residuals are homoscedastic and normally distributed; and 3) 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 & 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 test, 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.

2.1.1 The Common Factor Distribution

A skewed common factor distribution will result in skewed observed variables. A small number of studies 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 (Vermunt, 2004, Vermunt & Hagenaars, 2004, Muthén & Muthén, 2007) 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.

2.1.2 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 & 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 & Dolan (2009) presented a test of heteroscedastic residuals based on MML.

2.1.3 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., the factor loadings differ randomly between individuals. This results in a distribution of the observed variables that departs from normality (Kelderman & Molenaar, 2007; their plots suggest that random factor loadings tend to affect the kurtosis of the observed variable distribution). Ansari, Jedidi & 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 Chapter 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 denote this *level dependency* of the factor loadings. In the case that 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 &

Molenaar (2007) and Ansari et al. (2002) the randomness has 1) different consequences for the distribution of the observed data (it affects the skewness instead of the kurtosis), 2) it can be explained by taking η into account, and 3) 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., McDonald, 1962, Etezadi-Amoli & McDonald, 1983; Kenny & Judd, 1984; Jaccard & Wan, 1995; Klein & Moosbrugger, 2000; Lee & Zhu, 2002; Yalcin & Amemiya, 2001) . We elaborate on this below.

2.1.4 Motivation and Outline

This Chapter focuses on these three effects within the one factor model: non-normal factor distribution, heteroscedastic residuals, and level dependent factor loadings. Based partly on Hessen & Dolan (2009), we present a unified approach based on MML. This approach is promising, because: 1) it enables the test of the distributional assumption underlying factor analysis, i.e., that of a multivariate normal distribution of the observed data; 2) it enables the identification of specific violations of the multivariate normal distribution; and 3) once identified, it allows one to model these violations. The detection of specific violations within 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, Egan, Gibson, Austin, Brand & Kellaghan, 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). This approach extends the univariate manifest regression model, where the mean of the dependent variable is modeled as a function of the independent variables, by modeling the other parameters of the distribution of the dependent variable as well. In the present Chapter 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 therefore touches the approach of Rigby & Stasinopoulos, but is unique in the sense that it concerns an unobserved independent variable.

The organization of this Chapter 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. Thereafter, we fit the model to IQ data to illustrate the usefulness of the model in investigating ability differentiation. We conclude the Chapter with a general discussion.

2.2 The Formal Model and Estimation of the Parameters

2.2.1 Heteroscedastic Residuals

Hessen & 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 (Meijer & Mooijaart, 1996 and Lewin-Koh & Amemiya, 2003). 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 below, Equation 2). Here we generalize this approach. As in Hessen & Dolan (2009) heteroscedasticity is taken into account by modeling the logarithm of the residual variances conditional on η (see also Harvey, 1976). The conditional residual variance of variable j is then given by

$$\log(\sigma_{\varepsilon j}{}^{2} \mid \eta) = \beta_{jo} + \beta_{j1}\eta + \beta_{j2}\eta^{2} + ...\beta_{jr}\eta^{r}, \qquad (2)$$

i.e., the residual variance of variable j conditional on η, $\sigma_{\varepsilon j}^{2}|\eta$, is modeled by an r-th 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 $s = 1, ..., r$ 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 modeled to any degree of accuracy depending on the choice for r. For $r = 1$, a special model arises, which Hessen & Dolan refer to as the minimal heteroscedastic model, i.e.,

$$\log(\sigma_{\varepsilon j}{}^{2} \mid \eta) = \beta_{jo} + \beta_{j1}\eta, \qquad (3)$$

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

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

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

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

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

2.2.2 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 (Vermunt, 2004, Vermunt & Hagenaars, 2004, Muthén & Muthén, 2007). 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 freely, which are normally fixed to certain values (Schmitt, Mehta, Aggen, Kubarych, Neale, 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 Chapter 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

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

where ω is a scale parameter, κ is a location parameter, ζ is a shape parameter, φ(.) refers to the standard normal density function, and Φ(.) 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, 1990)

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

and

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

where

$$\delta = \frac{\zeta}{\sqrt{1+\zeta^2}} \qquad (8)$$

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

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

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

Equation 5 with Equation 6 and 7 form the skew-normal density as we use it here. Although there exist some literature concerning the generating process behind the skew-normal distribution (i.e., the density arises from a selection process; see Fernando de Helguero, 1908[1] and 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 test on normality straightforward by testing

$$H_0 : \zeta = 0.$$

[1] As cited by A. Azzalini at his website, see http://azzalini.stat.unipd.it/SN/ under the header 'A pioneer'

In addition, statistical properties of the skew-normal distribution are well documented (e.g., Azzalini & Dalla Valle, 1996; Azzalini & Capatanio, 1999; Arnold, Beaver, Groeneveld & Meeker, 1993; Arnold & Beaver, 2002; Chiogna, 2005; Monti 2003), the distribution is widely applied (for good examples, see Azzalini, 2005), and within 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.

2.2.3 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 modeling 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.,

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

which is a polynomial of the r-th degree. In this function, γ_{0j} is the baseline factor loading, which is independent of η, and γ_{sj} ($s = 1, ..., r$) 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 a choice for $s(.)$. Other functions could be used as well as we demonstrate in the application section.

As in the model for heteroscedasticity, we propose a minimal level dependency model within 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 Eq.1 we obtain

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

which can be fit using non-linear factor analyses (e.g., Klein & Moosbrugger, 2000; Lee & Zhu, 2002; Yalcin & Amemiya, 2001; Bauer, 2005a). 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 below. Fourth, in modeling 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 (2005a) pointed out that when the true factor-to-indicator relation is quadratic and invariant across groups, testing for measurement invariance within 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 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.

2.2.4 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 Marginal Maximum Likelihood (MML) to fit this model (Bock & Aitkin, 1981; see also Hessen & Dolan, 2009, and van der Sluis, Dolan, Neale, Boomsma & Posthuma, 2006). The loglikelihood function of the model is

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

where

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

and

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

W_q and N_q are transformations of the Q number of ‘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}, $E(\eta)$, $Var(\eta)$, and ζ), and $f(.)$ is given by

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

See the Appendix for a derivation of the loglikelihood function. It is important to note that the loglikelihood 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 number of class proportions. N_q^{*} are the nodes that are symmetric and sum to 0, these therefore correspond to the Q number of 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., Equation 13 and Equation 14). We only estimate the parameters from this transformation, instead of all W_q and N_q. The loglikelihood function can also be formulated in terms of a mixture distribution.

Then, we have Q mixtures of $f(\mathbf{y}_i \mid \eta, \boldsymbol{\tau})$, 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, 2006), but also in Mplus (Muthén & Muthén, 2007).

2.2.5 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 $Var(\eta)$ 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). Or, 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 the simulation section for some details on the empirical identification.

An issue related to identification is the curvature of the loglikelihood function. First, it is known that the loglikelihood 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 loglikelihood is monotome increasing for $\zeta \rightarrow \infty$ (see Chiogna, 2005; Monti 2003). The first anomaly is overcome by assigning nonzero starting values to ζ, and by refitting the model with different starting values for this parameter to see whether it makes a differences. The second anomaly did not appear to be a major problem, as is demonstrated in the simulation section below. Only in 3 out of 1800 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.

2.3 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.

2.3.1 Simulation Study 1: Design

We simulated data according to the model, as 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 modeling. In this undertaking, we manipulated 1) the number of subjects (3 levels: 300, 400, or 500 subjects); 2) 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 $\beta_{j1} \neq 0$ & $\zeta \neq 0$)[2]; and 3) the size of the effects (3 levels: small, medium, large). These manipulations gave rise to 36 conditions (3 x 4 x 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, 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 $Var(\eta)$ 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 density (α_1) is calculated according to Azzalini (1985)

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

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),

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

0.2 (medium), or 0.3 (large). To get an idea about the effect size of the manipulations of γ_{j1} and β_{j1}, we depicted in Figure 2.1 how the residual variances and the factor loadings vary as a function of η. From this Figure it appears that in case of a small effect size, both the factor loadings and the residual variances increase only a bit across η, while in 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 (Satorra & Saris, 1985; Saris & Satorra, 1993). 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 Satorra & Saris (1985) and Saris & Satorra (1993). 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., 2006) by minimizing -2 times the loglikelihood 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 www.dylanmolenaar.nl.

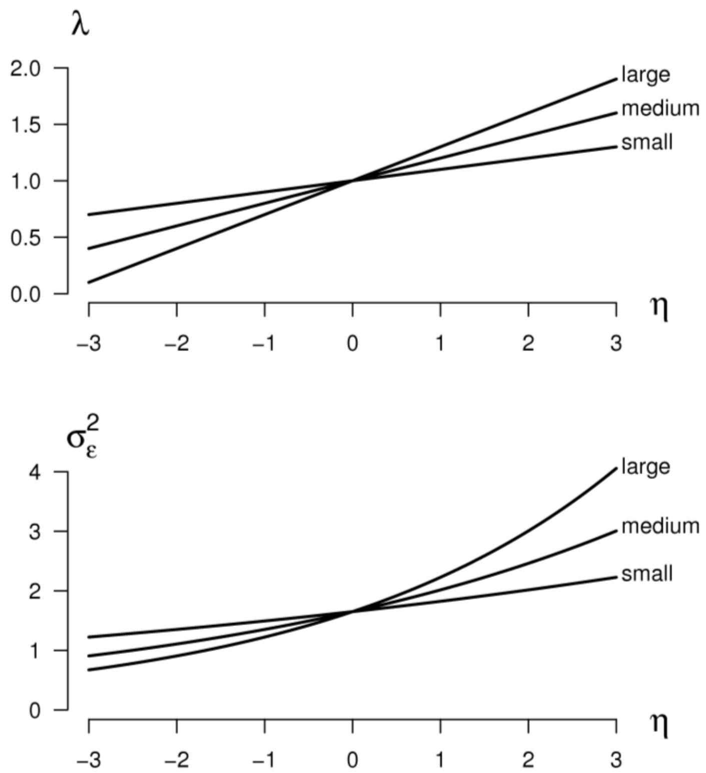

Figure 2.1. The factor loadings (top) and the residual variances (bottom) as a function of η for the small, medium, and large effect size condition.

2.3.2 Simulation Study 1: Results

In Table 2.1, the means and standard deviations of the estimates of the manipulated parameters are shown.[3] To ease presentation, we limited the Table to include only the results of items 1 to 3, N=300 and N=500, and a small and large effect. Judged by these results, parameter values are generally recovered quite well, and are quite unbiased.

In Table 2.2, rejection rates of the Shapiro-Wilks test on 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 datasets given an alpha of 0.01.

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 (0.01), while the results from Table 2.1 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 2.2 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:

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

As seen from Table 2.2, 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 2.2, the power the null-hypothesis that $\zeta = 0$ equals 0.99, which is large.

---

[3] 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 $\beta_{j1} \neq 0$ & $\zeta \neq 0$, N = 400 and a medium effect size. Both cases were removed from the subsequent results.

Table 2.1

Mean parameter estimates (and standard deviations) for items 1-3 and a small and large effect size in simulation study 1.

Effect	N	Model	ζ	γ_{11}	γ_{21}	γ_{31}	β_{11}	β_{21}	β_{31}
			1.81	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*	*0.1*
small	**300**	$\gamma_{j1} \neq 0$	-	0.09 (0.07)	0.10 (0.07)	0.09 (0.06)	-	-	-
		$\zeta \neq 0$	1.83 (0.73)	-	-	-	-	-	-
		$\gamma_{j1} \neq 0$ & $\beta_{j1} \neq 0$	-	0.10 (0.08)	0.11 (0.08)	0.10 (0.07)	0.11 (0.12)	0.10 (0.11)	0.12 (0.11)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	1.90 (0.71)	-	-	-	0.06 (0.07)	0.06 (0.08)	0.07 (0.07)
	500	$\gamma_{j1} \neq 0$	-	0.10 (0.06)	0.10 (0.05)	0.09 (0.06)	-	-	-
		$\zeta \neq 0$	1.75 (0.39)	-	-	-	-	-	-
		$\gamma_{j1} \neq 0$ & $\beta_{j1} \neq 0$	-	0.11 (0.06)	0.10 (0.06)	0.09 (0.05)	0.10 (0.08)	0.10 (0.09)	0.10 (0.09)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	1.82 (0.46)	-	-	-	0.06 (0.06)	0.08 (0.05)	0.07 (0.05)
			2.62	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*
large	**300**	$\gamma_{j1} \neq 0$	-	0.30 (0.10)	0.30 (0.11)	0.30 (0.10)	-	-	-
		$\zeta \neq 0$	2.91 (1.12)	-	-	-	-	-	-
		$\gamma_{j1} \neq 0$ & $\beta_{j1} \neq 0$	-	0.37 (0.21)	0.37 (0.18)	0.36 (0.20)	0.33 (0.14)	0.33 (0.14)	0.34 (0.15)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	2.80 (0.91)	-	-	-	0.20 (0.07)	0.19 (0.07)	0.20 (0.06)
	500	$\gamma_{j1} \neq 0$	-	0.30 (0.07)	0.30 (0.07)	0.30 (0.07)	-	-	-
		$\zeta \neq 0$	2.78 (0.74)	-	-	-	-	-	-
		$\gamma_{j1} \neq 0$ & $\beta_{j1} \neq 0$	-	0.31 (0.09)	0.31 (0.09)	0.30 (0.09)	0.29 (0.09)	0.31 (0.08)	0.32 (0.10)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	2.81 (0.67)	-	-	-	0.19 (0.05)	0.19 (0.05)	0.19 (0.05)

Note. True parameter values are in underlined. Dashes mean that the corresponding parameter is not in the model.

Table 2.2.

Rejection rates and empirical power coefficients for simulation study 1

Effect	**N**	**Model**	**Rejection rate item no**					**Power**	
			1	**2**	**3**	**4**	**5**	**1**	**1 & 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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	16	24	19	18	21	0.68 (10.11)	0.77 (12.01)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	27	25	31	23	34	0.77 (12.14)	0.83 (13.80)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	39	41	40	32	46	0.83 (13.62)	0.90 (16.24)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	71	75	76	71	76	0.98 (24.78)	0.99 (28.76)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	84	95	90	87	91	1.00 (30.99)	1.00 (37.84)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	95	97	95	93	99	1.00 (38.08)	1.00 (49.54)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	99	97	100	100	100	1.00 (51.23)	1.00 (62.92)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	100	100	100	100	100	1.00 (63.53)	1.00 (80.35)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	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)	-
		$\beta_{j1} \neq 0$ & $\gamma_{j1} \neq 0$	100	100	100	100	100	1.00 (79.92)	1.00 (94.91)
		$\zeta \neq 0$ & $\beta_{j1} \neq 0$	95	98	98	98	98	1.00 (20.64)	1.00 (111.40)

Note. 'Power 1': Power to detect the effect that is listed first (i.e., $\beta_{j1} \neq 0$ or $\zeta \neq 0$). 'Power 1&2': power to detect both effects simultaneously. Empirical non-centrality parameters are in brackets.

2.3.3 Simulation Study 2: Design

In this study we simulated 100 datasets according to each of three models: 1) a model with heteroscedastic residuals ($\beta_{j1} \neq 0$); 2) a model with level dependent factor loadings ($\gamma_{j1} \neq 0$); and 3) 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: 1) a model with $\beta_{j1} \neq 0$ and $\gamma_{j1} \neq 0$; and 2) 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.

2.3.4 Simulation Study 2: Results

In Table 2.3, means and standard deviations of the estimated parameters are shown[4]. 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.

2.4 Application

2.4.1 Background.

A well established phenomenon in intelligence research is that all subtest scores of an IQ test (e.g., the 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 toward the higher end of *g*

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

Table 2.3.

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}
ζ **= 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)
γ_{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)
β_{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.

(Spearman, 1927). The ability differentiation hypothesis has enjoyed a good deal of attention during the past decennia, which resulted in many papers (e.g., Detterman & Daniel, 1989; Deary, et al. 1996; Jensen, 2003; Facon, 2004, 2008; Hartmann & Teasdale, 2004; Abad, Colom, Juan-Espinosa & Garcia, 2003; Carlstedt, 2001, Reynolds & Keith, 2007; Fogarty & Stankov, 1995). 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), however this is hard to do as *g* is an unobserved variable. Reynolds & 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 (Saris, de Pijper, & Mulder, 1978; Jöreskog, Sörbom, du Toit, & du Toit, 1999). 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 be 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 & 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 al $j = 1, ..., p$. Finally, ability differentiation could be caused by level dependency in the factor loadings, with smaller factor loadings at the higher levels of *g*, i.e., $\gamma_{j1} < 0$ for all $j = 1, .., p$. 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 be the cause.

2.4.2 Analysis and Results

In this application, we specify different functions between λ and η within 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. Disadvantage of this function is that for large values of η relative to γ_{j0}, the $\lambda_j(\eta)$ assume negative value. In modelling intelligence, we want to avoid this so as to retain the positive manifold at all levels of λ_j.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.,

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

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, within 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 Chapter (e.g., Equation 2 and Equation 9).

We analyzed data from the National Longitudinal Survey of Youth 1997[5]. From the 12 subtests of the Armed Services Vocational Aptitude Battery (ASVAB), we selected 5 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 Table 2.4 and 2.5.

Table 2.4.
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

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

Table 2.5.
Correlations among the variables in the application.

	AR	PC	AI	MK
PC	0.73			
AI	0.53	0.49		
MK	0.80	0.75	0.45	
AO	0.60	0.58	0.38	0.59

The total sample size consisted of 7127 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.e-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 by a 0.01 significance level (AR, W = 0.99, $p < 0.001$; PC, W = 0.98, $p < 0.001$; AI, W = 0.99, p = 0.003; MK, W = 0.99, p = 0.028; AO, W = 0.98, $p < 0.001$). We first fitted the standard one factor model, as a baseline model. The model was identified by setting γ_{10} of MK to equal 1[6], and $E(\eta)$ to equal 0. The fit of the model was acceptable (-2 *logL* = 6188.91, df = 2467, 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 2.6 for β_{j1}. As appears from these results, for the subtests AR and OA 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 2.6 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 OA as indicated by the significant heteroscedaticity parameter estimates, β_{j1}, for these subtests.

[6] In case of the exponential model for the factor loadings, γ_{10} was fixed to log(1) to facilitate interpretation across models.

Table 2.6.
Parameter estimates and likelihood ratio test statistics for the application.

Model	LRT*	df		AR	PC	AI	MK	AO
baseline			γ_{j1}	-	-	-	-	-
			β_{j1}	-	-	-	-	-
$\beta_{j1} \neq 0$	35.16	5	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)
$\gamma_{j1} \neq 0$ &	39.54	10	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)
$\gamma_{j1} \neq 0$ &	39.51	10	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)
$\gamma_{j1} \neq 0$ &	23.85	10	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)
$\zeta \neq 0$ &	34.45	6	γ_{j1}	-	-	-	-	-
$\beta_{j1} \neq 0$								
			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 logL* between that model and the baseline model. For the parameter estimate of ζ in the model at the bottom of the Table, see the text.

However, the heteroscedasticity can not be reasonably interpreted in terms of ability differentiation, as AR has a negative estimate of β_{j1}, indicating that the residual variance decreases along the g (or η) range, while AO has a positive estimate β_{j1}, indicating that the residual variance of this subtest increases along 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 of non-significance[7]. To avoid possible local minima, we used different starting values for the shape parameter, ζ. In the model, ζ was estimated to be -0.71 which is not significant according to its confidence interval (99% confidence interval bounds equalled -2.32 and 1.30). The overall conclusion is therefore that there is no level

[7] 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.

dependency in the factor loadings or skewness in the factor distribution. In addition, 2 subtests showed heteroscedastic residuals, but the direction of these effects were not in line with the hypothesized differentiation effects.

2.5 Discussion

We presented a unified model to test the distributional assumptions underlying Maximum Likelihood estimation within the one factor model. In addition to testing, the model is useful in modeling 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 modeled simultaneously.

Testing and modeling violations of distributional assumptions within any statistical model may be interesting from theoretical considerations. To illustrate this, we discussed how an empirical phenomenon, ability differentiation, implies specific violations within the common factor model. Once identified, the presence of the phenomenon was easily tested in a dataset 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 underlied 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 Chapter, we wish only to stress that 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 can not be estimated simultaneously. Nonetheless, we think that both are valuable in their own rights. 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 multi-dimensional. 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, Wilson, Gibbons, Schilling, Muraki & Bock, 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 Chapter could be easily 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 2 Parameter Logistic model with level dependent discrimination parameters. This possibility has yet to be pursued.

In this Chapter we limited the simulations and application to only 5 variables. Consequently, the question arises how the model performs with more variables. From a numerical point of few adding variables merely increases computational 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 one 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 exist that the residual covariances vary across the common factor as well. In the Mx syntax on www.dylanmolenaar.nl, 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 along 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 sufficient sample sizes, 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 bit (up to 10%) but still, the majority of the cases converged at reasonably values. Note that these sample size indications depends 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 that are modeling psychological phenomenon.

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 departure 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.

Appendix: Derivation of the loglikelihood function of the model.

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

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

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 $\sigma_{\varepsilon j}^{2} \mid \eta$ and $\lambda_j(\eta)$ from respectively Equation 2 and Equation 9 resulting in

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

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

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

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

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

Recap that φ(.) and Φ(.) 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

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

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

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

We transform η using the transformation

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

with

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

This results in

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

This expression could be approximated with Gauss-Hermite quadratures resulting in

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

By specifying

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

and

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

the expression reduces to

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

which is in de form of Equation A4.

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

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

A final step is to take the logarithm of Equation A6 to obtain the marginal loglikelihood function,

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