The British
Psychological Society

Special Issue Paper

A heteroscedastic generalized linear model with a non-normal speed factor for responses and response times

Dylan Molenaar* and Maria Bolsinova

University of Amsterdam, The Netherlands

In generalized linear modelling of responses and response times, the observed response time variables are commonly transformed to make their distribution approximately normal. A normal distribution for the transformed response times is desirable as it justifies the linearity and homoscedasticity assumptions in the underlying linear model. Past research has, however, shown that the transformed response times are not always normal. Models have been developed to accommodate this violation. In the present study, we propose a modelling approach for responses and response times to test and model non-normality in the transformed response times. Most importantly, we distinguish between non-normality due to heteroscedastic residual variances, and non-normality due to a skewed speed factor. In a simulation study, we establish parameter recovery and the power to separate both effects. In addition, we apply the model to a real data set.

1. Introduction

One of the dominant approaches to the analysis of responses and response times is the generalized linear modelling framework (GLM; Molenaar, Tuerlinckx, & Van der Maas, 2015a; Ranger, 2013; Ranger & Ortner, 2012a). In this framework, a measurement model is specified for the responses to operationalize a normally distributed latent ability factor. For the response times, a measurement model is specified to operationalize a normally distributed latent speed factor. Special cases within the GLM differ in the exact way in which both measurement models are connected. By specifying a linear relation between the ability factor and the response times, the hierarchical models by Van der Linden (2007, 2009) and Thissen (1983; see also Furneaux, 1961) arise. By using a non-linear relation between the latent ability factor and the response times, the distance–difficulty models for personality data arise (Ferrando & Lorenzo-Seva, 2007a,b).

As response times are typically bounded by zero and skewed (Luce, 1986), researchers have transformed the raw response times to enable linear modelling of them. Such transformations mainly serve the purpose of linearizing the relation between the response times and the underlying normally distributed speed factor, and of making the assumption

*Correspondence should be addressed to Dylan Molenaar, Psychological Methods, Department of Psychology, University of Amsterdam, Postbus 15906, 1001 NK Amsterdam, The Netherlands (email: D.Molenaar@uva.nl).

of homoscedasticity of the residual variance in the linear response time model more plausible. Transformations that have been considered include the reciprocal (e.g., Ferrando & Lorenzo-Seva, 2007b), square root (e.g., Aberson, Shoemaker, & Tomolillo, 2004), and logarithm (e.g., Thissen, 1983; Van der Linden, 2007). Assuming that the transformed response times are indeed normal, these transformations imply respectively an exponential, chi-square, and log-normal distribution for the untransformed response times.

The log transformation is arguably the most widely used. However, as discussed by Klein Entink, Linden, and Fox (2009) and Ranger and Kuhn (2012), the implied log-normal distribution does not always provide a satisfactory approximation for the observed response time distributions. To this end, Klein Entink *et al.* proposed a more flexible Box–Cox normal model for the response times. In these models, an extra parameter is estimated which reflects the degree of non-normality in the log-response times. In addition, various semi-parametric approaches have been proposed, including a proportional hazards model (Loeys, Legrand, Schettino, & Pourtois, 2014; Ranger & Ortner, 2012b, 2013; Wang, Fan, Chang, & Douglas, 2013), a linear transformation model (Wang, Chang, & Douglas, 2013), and models for categorized response times (Ranger & Kuhn, 2012, 2013).

Here, a new approach is outlined to account for departures from normality in the transformed response times. This approach is derived by noting that, in the linear model for the transformed response times,

$$T_{pi}^* = \nu_i - \tau_p + \varepsilon_{pi},$$

normality of the transformed responses times, T_{pi}^*, implies a normal speed factor, τ_p, and homoscedastic and normal residuals, ε_{pi}. That is, departures from normality in T_{pi}^* will be solely due to (1) non-normality in τ_p, (2) non-normality in ε_{pi} and/or (3) heteroscedastic ε_{pi}. The approach by Klein Entink *et al.* (2009) focuses on (2). In the present paper, we provide a test on normality of T_{pi}^* based on (1) and (3). That is, we explicitly separate between non-normality due to a non-normal latent speed factor and between non-normality due to heteroscedastic residuals.

1.1 Non-normal speed factor

A non-normal speed factor may arise if there are qualitative between-subject differences in response speed. For instance, Meyer (2010) showed that if some subjects resort to rapid guessing for all items in a test, the transformed response time distribution will have a longer lower tail which can be captured by a two-class between-subject mixture model. Although Meyer estimated the mixture for each item separately, a more parsimonious way would be to introduce the mixture into the speed factor as a rapid guessing subject will have an overall higher speed level on all items as compared to a non-guessing subject. The existence of such a between-subject mixture will thus show up as non-normality in the speed factor in the present approach. Another example of a between-subject mixture of response speed is given by Van der Maas and Jansen (2003) who showed that children on a balance scale task adopt different solution strategies which differ in their execution time. Again, such heterogeneity causes the distribution of the speed factor to depart from normality.

1.2. Heteroscedastic residuals

In this paper, we use the term 'heteroscedasticity' to refer to the observation that the residual variances in a linear model are not constant across the predictors. In linear

regression models, heteroscedasticity is a well-studied phenomenon. Methods have been proposed to test the equality of the residual covariance matrix across the levels of the predictor variables (e.g., Anderson, 2006), and methods have been studied to account for possible violations (e.g., Brunner, Dette, & Munk, 1997). In addition, various approaches exist to assess, test or model heteroscedasticity in linear regression models, including diagnostic graphical approaches (e.g., Stevens, 2009, p. 90), statistical tests (e.g., Jarque & Bera, 1980), corrections (e.g., Long & Ervin, 2000), and approaches to model heteroscedasticity explicitly (e.g., Harvey, 1976). Interestingly, although the GLM for responses and response times is in essence also a linear regression model (but with a single latent predictor variable), tests on heteroscedasticity have not yet received attention. In GLMs for responses and response times, heteroscedasticity can arise for various reasons, including item deadlines (censoring), the use of an inappropriate transformation function for the untransformed response times, and the existence of item-specific mixtures. Examples of item-specific mixtures include item pre-knowledge (McLeod, Lewis, & Thissen, 2003), post-error slowing (Rabbitt, 1979), and rapid guessing (Wang & Xu, 2015) on some of the items but not on all.

1.3. Motivation

In the present paper, we will present a unified model that enables the simultaneous estimation of heteroscedasticity in the residuals and non-normality in the speed factor. Such an endeavour is valuable for a number of reasons:

Explicit statistical test. The assumption of normality can be tested using marginal tests like the test of Mardia (1970) on multivariate normality and that of Shapiro and Wilks (1965) on univariate normality. Such tests are marginal tests as they do not impose a latent variable structure on the data. We will show in the present study that by using a latent variable modelling framework for the responses and response times, the power to detect an effect is greater than the multivariate test by Mardia (1970). An alternative approach to the marginal tests, used by for instance Van der Linden, Breithaupt, Chuah, and Zhang (2007) and Van der Linden, Scrams, and Schnipke (1999) employs quantile–quantile plots. As discussed by Klein Entink *et al.* (2009), although valuable, quantile–quantile plots may be hard to interpret and they do not penalize for model complexity. Therefore, an explicit statistical test may be desirable.

Affected person parameter estimates. As we will illustrate in this paper, results from response and response time analyses may be affected by the presence of non-normality in the transformed response time distribution. That is, we find that non-normality affects the ability and speed estimates in the tails of the latent ability and speed factor distributions. Although such a discrepancy may not always arise, it is valuable to have a tool at our disposal that can be used to test the normality assumption and compare the results across the normal and non-normal models. If results do not differ importantly, one can safely rely on the normal modelling results. However, if results differ (as is the case in our illustration below), one may better rely on the results from the best-fitting model, as these are likely to capture the patterns in the data better.

Aberrant responses. The latent ability and latent speed factor estimates play an important role in the detection of aberrant responses (Marianti, Fox, Avetisyan, Veldkamp, & Tijmstra, 2014; Van der Linden & Guo, 2008; Van der Linden & van Krimpen-Stoop, 2003) and in the selection of items in computerized adaptive testing (Van der Linden *et al.*, 1999). As discussed above, for the respondents in the tails of the ability and speed distribution, the speed and ability estimates may be affected by non-normality

in the transformed response times. Therefore, for these respondents, it is important that the most appropriate estimates are used to prevent invalid inferences about aberrances or to prevent a suboptimal item selection procedure in the computerized adaptive testing procedure. In the case of severe non-normality in the transformed response time distribution, the ability and speed factor estimates from the non-normal models might thus be preferred.

Using an appropriate transformation. Another motivation for having a statistical test on normality available within the model for responses and transformed response times is that it is often unclear which transformation (e.g., logarithmic, reciprocal, or square root) to adopt for the response times. That is, it is unclear what transformation makes the linearity and homoscedasticity assumptions plausible. Using the present approach, it can be investigated which transformation works best in normalizing the transformed response times. If normality does not hold for any of the transformations considered, the resulting violation of homoscedasticity and linearity can be explicitly taken into account in the model.

Increasing power. In the present paper, the reason to focus on both heteroscedasticity and non-normality in the speed factor is mainly to increase the resolution to detect departures from normality. As we will show in this paper, both effects are statistically well separable, meaning that the effects of a non-normal speed factor can hardly be captured by heteroscedastic residuals, and vice versa. Therefore, as both effects pick up distinct patterns in the data, taking both effects into account increases the power to detect non-normality.

Personality items. Although the main focus of this paper is on ability measurement, some of the above motivations also hold for applications to personality items. That is, faking on personality items (Holden & Kroner, 1992) may be detected using the same procedures for aberrant responses as discussed above. In addition, research into responses and response times in personality have focused on the distance–difficulty hypothesis (Ferrando & Lorenzo-Seva, 2007a,b) or the inverted-U response time hypothesis (Kuiper, 1981). This hypothesis states that for personality items, if the distance between the item difficulty parameters and the latent ability parameters increases, response times decrease. As the models used to test this effect include a quadratic (or approximately quadratic) effect of the latent ability factor on the transformed response times (see Molenaar, Tuerlinckx, & van der Maas, 2015b), the transformed response times will be non-normally distributed if the distance–difficulty hypothesis holds. The models put forward in this paper may thus provide an omnibus test for the distance–difficulty hypothesis from personality research.

1.4. Outline

The outline of this paper is as follows. First, the GLM approach to responses and response time modelling is outlined. Next, we introduce heteroscedastic residuals and a non-linear speed factor. In a simulation study we show that the model is viable in terms of parameter recovery and the power to separate between different sources of non-normality in the transformed response times. We then apply the model to a real data set and illustrate that the latent ability and latent speed factor estimates may be different in the normal and non-normal models for respondents in the tails of the distributions. We conclude with a general discussion.

2. The generalized linear model for responses and response times

Let X_{pi} denote the response of subject p on item i with transformed response time T_{pi}^*. As discussed above in the GLM framework, a measurement model is specified for both variables. To connect the two models, Van der Linden (2007, 2009) and Glas and van der Linden (2010) used a common distribution for the item parameters and a common distribution for the person parameters in the two models. In addition, Molenaar *et al.* (2015b), Ranger and Ortner (2012a), Van der Linden and Guo (2008), Wang, Change, *et al.* (2013), and Wang, Fan *et al.* (2013) specified a common distribution for the person parameters only. Here we follow Molenaar *et al.* (2015a), Ranger (2013), Thissen (1983), Furneaux (1961), and Ferrando and Lorenzo-Seva (2007a,b) and specify a cross-loading of the response times on the latent ability factor. That is, by assuming independence of X_{pi} and T_{pi}^* conditional on the random person parameters, the joint likelihood factors as

$$l(X_{pi}, T_{pi}^*|\theta_p, \tau_p) = f(X_{pi}|\theta_p)g(T_{pi}^*|\tau_p, \theta_p)\phi(\theta_p)\phi(\tau_p), \tag{1}$$

where $f(\cdot)$ is the likelihood function according to the measurement model of the responses, $g(\cdot)$ is the likelihood function according to the measurement model of the transformed response times including the cross-loading on θ_p, and $\phi(\cdot)$ is a standard normal density function for the latent ability parameters, $\theta_p \in (-\infty, \infty)$, and the latent speed parameters, $\tau_p \in (-\infty, \infty)$, respectively. We elaborate on choices for $f(\cdot)$ and $g(\cdot)$ below.

For $f(\cdot)$ in equation (1), the measurement model of the responses, researchers have used the Rasch model (Loeys *et al.*, 2014), the two-parameter model (Molenaar *et al.*, 2015a,b; Ranger & Ortner, 2012a; Thissen, 1983), the graded response model (Ranger, 2013) and the linear factor model (Ferrando & Lorenzo-Seva, 2007b). Here we use the two-parameter model, but we note that present undertaking is equally amenable to other generalized linear models (e.g., a graded response models in the case of Likert scale personality item scores). We thus specify

$$f(X_{pi}|\theta_p) = \frac{\exp[X_{pi}(\alpha_i\theta_p - \beta_i)]}{1 + \exp(\alpha_i\theta_p - \beta_i)}, \tag{2}$$

where $\alpha_i \in (-\infty, \infty)$ is the discrimination parameter which is expected to be strictly positive in practice and $\beta_i \in (-\infty, \infty)$ is the difficulty parameter.

For $g(\cdot)$ in equation (1), the measurement model of the response times including the cross-loading on θ_p, researchers have mainly used a normal model. However, others have used a proportional hazards model (Loeys *et al.*, 2014; Wang, Fan *et al.* 2013), a linear transformation model (Wang, Change, *et al.*, 2013), or a categorical model for discretized time (Ranger & Kuhn, 2012, 2013). There are many possible parameterizations of the normal model. Here we focus on

$$g(T_{pi}^*|\tau_p, \theta_p) = \frac{1}{\sqrt{2\pi\sigma_{\varepsilon i}^2}}\exp\left\{-\frac{1}{2}\frac{[T_{pi}^* - (\nu_i - \lambda_\theta\theta_p - \lambda_\tau\tau_p)]^2}{\sigma_{\varepsilon i}^2}\right\}, \tag{3}$$

where $\nu_i \in (-\infty, \infty)$ is the time intensity parameter, and $\sigma_{\varepsilon i}^2 \in [0, \infty)$ is the residual variance, that is, the variance of T_{pi}^* conditional on τ_p. In equation (3), $\lambda_\theta \in (-\infty, \infty)$ denotes the cross-loading of the response times on the ability factor θ_p to connect the

measurement model of the responses to the measurement model of the response times (see Furneaux, 1961; Molenaar *et al.*, 2015a; Ranger, 2013; Thissen, 1983). In addition, in equation (3), $\lambda_\tau \in (-\infty, \infty)$ models the scale of τ_p. Note that the model in equation (3) is equivalent to the hierarchical model of Van der Linden (2007, 2009) with fixed item effect (see Molenaar *et al.*, 2015a; Ranger, 2013). That is, the correlation between τ_p and θ_p in the hierarchical model, ρ, can be retrieved from the parameters in equation (3) as follows:

$$\rho = \frac{\lambda_\theta}{\sqrt{\lambda_\theta^2 + \lambda_\tau^2}}. \tag{4}$$

In addition, the variance of τ_p in the hierarchical model, σ_τ^2, can be retrieved from the parameters in equation (3) as follows:

$$\sigma_\tau^2 = \frac{\lambda_{\theta^2}}{\rho^2} = \lambda_\theta^2 + \lambda_\tau^2. \tag{5}$$

Note that the cross-loading from the present parameterization, λ_θ, is an unscaled version of the correlation parameter, ρ, from the hierarchical model. As the correlation parameter does not depend on the item, λ_θ does not depend on the item either.

3. A heteroscedastic generalized linear model with a skewed speed factor

Here, we extend the traditional model in equations (1–3) by allowing non-normality in the speed factor and by allowing the residual variances to be heteroscedastic.

3.1. Heteroscedastic residuals

In factor analysis, methods to account or test for heteroscedasticity include the generalized least squares procedure of Meijer and Mooijaart (1996), the distribution-free method of Lewin-Koh and Amemiya (2003), and the two-stage least squares procedure of Bollen (1996). Here, we adopt the approach of Hessen and Dolan (2009) which is developed within the GLM framework as adopted in the present paper. Specifically, Hessen and Dolan proposed to model heteroscedasticity within the linear one-factor model by modelling the variance of the indicators conditional on the common factor, $\sigma^2|\eta_p$, by an exponential function (for a similar approach within regression modelling, see Harvey, 1976). Specifically,

$$\sigma^2|\eta_p = \exp(\delta_{0i} + \delta_{1i}\eta_p), \tag{6}$$

where η_p is the common factor, $\delta_{0i} \in (-\infty, \infty)$ is an intercept parameter and $\delta_{1i} \in (-\infty, \infty)$ is a heteroscedasticity parameter which accounts for the dependency of the conditional variance on η. That is, for $\delta_{1i} = 0$, the residual variance is homoscedastic, for $\delta_{1i} > 0$, the residual variance is increasing for increasing levels of η_p, and for $\delta_{1i} < 0$, the residual variance is decreasing for increasing levels of η_p.

To introduce heteroscedastic residuals into the traditional GLM given by equation (3), the residual variances are explicitly conditioned on τ_p and modelled as in equation (6), where $\eta_p = -\lambda_\tau\tau_p$. This results in

$$g(T_{pi}^*|\tau_p) = \frac{1}{\sqrt{2\pi \times \exp(\delta_{0i} - \delta_{1i}\lambda_\tau\tau_p)}} \exp\left\{-\frac{1}{2}\frac{[T_{pi}^* - (\nu_i - \lambda_\theta\theta_p - \lambda_\tau\tau_p)]^2}{\exp(\delta_{0i} - \delta_{1i}\lambda_\tau\tau_p)}\right\}. \quad (7)$$

Note that we explicitly focus on heteroscedasticity across τ_p and not across θ_p.

3.2. Non-normal speed factor

Various approaches exist by which non-normal latent variables can be modelled. For instance, the latent distribution can be approximated using a constrained mixture approach (Schmitt, Mehta, Aggen, Kubarych, & Neale, 2006; Vermunt, 2004; Vermunt & Hagenaars, 2004), Johnson curves (Van den Oord, 2005), and splines (Woods, 2007). In addition, researchers have used the log-beta distribution (Andersen & Madsen, 1977). Here, we use the skew normal distribution (Azzalini, 1985, 1986). The density of the skew normal distribution is given by

$$k(\eta) = \frac{2}{\sqrt{2\pi\omega^2}}\Phi(\zeta\frac{\eta - \kappa}{\omega})\exp(-\frac{1}{2}\frac{(\eta - \kappa)^2}{\omega^2}), \quad (8)$$

where $\kappa \in (-\infty, \infty)$ is a location parameter, $\omega^2 \in [0, \infty)$ is a scale parameter, and $\zeta \in (-\infty, \infty)$ is a skewness parameter. That is, for $\zeta = 0$ the density above reduces to the density of a normal distribution, for $\zeta > 0$ the density is right-skewed, and for $\zeta < 0$ the density is left-skewed.

Our choice of the skew normal distribution is driven by the following considerations:

1. As the skew normal density includes the normal distribution as special case for $\zeta = 0$, the skew normal density enables a straightforward likelihood ratio test on normality with only one degree of freedom. If we had adopted a two-class mixture distribution (which is certainly feasible), we would need test with at least two degrees of freedom (the mixing proportion and the mean of the second mixture component), probably lowering power. In addition, a likelihood ratio test is not feasible as a mixture distribution would require boundary constraints to arrive at a normal distribution. As the statistical properties of the likelihood ratio are well known, we considered this test desirable. However, a mixture distribution has some advantages over the skew normal distribution. We return to this point in Section 6.
2. The skew normal density is unbounded, which is appropriate for the unbounded transformed response times. The beta distribution and the log-normal distribution, for instance, are bounded and thus less suitable for this endeavour.
3. The skew normal density is well developed within the present generalized linear framework (see Azevedo, Bolfarine, & Andrade, 2011; Molenaar, Dolan, & de Boeck, 2012).
4. A linear factor model with a skew normal distribution for the factor is shown to be equivalent to a one-factor model with quadratic factor loadings (Smits, Timmerman, & Stegeman, 2015). As such non-linear factor models are well developed (e.g., Klein & Moosbrugger, 2000; McDonald, 1962), we can draw from the modelling tools available for these models and use them in our modelling approach (e.g., parameter estimation, factor score estimation). In addition, due to this equivalence, our modelling approach can be implemented in existing software, hopefully increasing the practical utility of the approach.

To introduce a skew normal distribution for τ_p, we use the result of Smits *et al.* (2015) mentioned above. Specifically, Smits *et al.* show that a one-factor model with a skew normal factor distribution given by equation (8) is equivalent to a model with quadratic factor loadings up to the first three sample moments.[1]

Here, we use this equivalence to specify the skew normal speed factor as a quadratic speed factor. That is, in the model subject to heteroscedasticity in equation (7) we additionally introduce the effect of non-normality in τ_p by adding the quadratic effect of τ_p, that is,

$$g(T_{pi}^*|\tau_p, \theta_p) = \frac{1}{\sqrt{2\pi \exp(\delta_{0i} - \delta_{1i}[\lambda_\tau \tau_p + \lambda_\zeta \tau_p^2])}} \exp\left\{ -\frac{1}{2} \frac{\left[T_{pi}^* - (\nu_i - \lambda_\theta \theta_p - [\lambda_\tau \tau_p + \lambda_\zeta \tau_p^2]) \right]^2}{\exp(\delta_{0i} - \delta_{1i}[\lambda_\tau \tau_p + \lambda_\zeta \tau_p^2])} \right\}, \tag{9}$$

where $\lambda_\zeta \in (-\theta, \theta)$ is a shape parameter that corresponds to ζ in the skew normal density in equation (8). See Smits *et al.* for the exact relation between ζ and λ_ζ. Note that the equivalence between the quadratic model in equation (9) and a model with a skew normal τ_p specified through the density function of τ_p in equation (1) only holds for the first three sample moments. However, Smits *et al.* show by means of the L_1 norm of the difference in the model-implied data distributions that for values of ζ of practical interest, the differences in the two implied distributions are only minor if all moments are considered (i.e., as is the case with full-information procedures as adopted here in equation (1)).

3.3. Estimation of the parameters

Together with the two-parameter model for X_{pi} the marginal log-likelihood of the matrix of item responses, $\mathbf{X}$, and the matrix of transformed response times, $\mathbf{T}^*$, given the model parameters, $\boldsymbol{\omega} = [\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n, \nu_1, \ldots, \nu_n, \delta_{01}, \ldots, \delta_{0n}, \delta_{11}, \ldots, \delta_{1n}, \lambda_\theta, \lambda_\tau, \lambda_\zeta]$ is given by

$$\ell(\mathbf{X}, \mathbf{T}^*; \boldsymbol{\omega}) = \sum_{p=1}^N \ln \int \int_{-\infty}^{\infty} \prod_{i=1}^n f(X_{pi}|\theta_p) g(T_{pi}^*|\tau_p, \theta_p) \phi(\theta_p) \phi(\tau_p) \mathrm{d}\theta \mathrm{d}\tau, \tag{10}$$

where n denotes the number of items, N denotes the number of subjects, $f(\cdot)$ is given by equation (2) and $g(\cdot)$ is given by equation (9).

All free parameters in the likelihood function in equation (10) can be estimated simultaneously using marginal maximum likelihood (MML) estimation (Bock & Aitkin, 1981). To this end, the integrals in the likelihood function are approximated by Gauss–Hermite quadrature (GHQ). We implemented the model in Mplus (Muthén & Muthén, 2007) using 15 nodes for each of the θ_p and τ_p dimensions. As the

---

[1] Specifically, let $T = a + bX + e$ denote a factor model with the factor, X, subject to a skew normal distribution with skewness parameter s. Next, let $T' = a' + cX' + dX'^2$ denote a quadratic factor model with the factor, X', subject to a normal distribution. Smits *et al.* show that you can express d (the quadratic loading effect) in terms of s (the skewness parameter) such that T and T' are equivalent up to the first three moments.

syntax to fit the model can become very large for an increasing number of items, n, we implemented an R script that generates the Mplus syntax to fit the model for any choice for n and number of nodes. In addition, we wrote some documentation which gives a guide to how to read the Mplus output. See http://www.dylanmolenaar.nl for the R script and the documentation.

3.4. Evaluating model fit

As for $\delta_{1i} = 0$ the residuals are homoscedastic and for $\lambda_\zeta = 0$ the speed factor is normal, inferences about the sources of non-normality in the transformed response times can be based on models that do and do not include these parameters. Note that as both constraints do not involve a boundary constraint, a likelihood ratio test or fit indices like the Akaike information criterion (AIC; Akaike, 1974), Bayesian Information Criterion (BIC; Schwarz, 1978), and the saBIC (sample size adjusted BIC; Sclove, 1987) are suitable for model comparisons with and without heteroscedasticity and/or a skewed latent speed factor. However, the standard errors of these parameters might be biased as, within MML estimation, these are based on the assumption of an asymptotic normal parameter distribution which is likely to be violated for the parameter distributions of δ_{1i} and λ_ζ.

4. Simulation study

4.1. Design

To investigate parameter recovery, Type I error, and the power to distinguish between the different sources of non-normality in the transformed response times, a simulation study was conducted. In this study we focus on four models:

1. *M1: het-GLM-skew*. A heteroscedastic generalized linear model for responses and response times with a skew normal speed factor where λ_ζ is a free parameter and δ_{1i} are free parameters for all i.
2. *M2: het-GLM*. A heteroscedastic generalized linear model for responses and response times with a normal speed factor where $\lambda_\zeta = 0$ and δ_{1i} are free parameters for all i.
3. *M3: GLM-skew*. A homoscedastic generalized linear model for responses and response times with a skew normal speed factor where λ_ζ is a free parameter and $\delta_{1i} = 0$ for all i.
4. *M4: GLM*. A homoscedastic generalized linear model for responses and response times with a normal speed factor where $\lambda_\zeta = 0$ and $\delta_{1i} = 0$ for all i (i.e., the traditional model given by equations (1–3)).

We simulated data according to these four models. In this simulation study, we used the log transformation for the response times, that is, $T_{pi}^* = \ln(T_{pi})$. In addition, we used $N = 500$ and $n = 20$. Difficulty parameters, β_i, were chosen at equally spaced values in the $[-3, 3]$ interval. Discrimination parameters, α_i, were equal to 0.5 and 1.5 for the odd and even items, respectively. Time intensity parameters, ν_i, were equal to 2. In addition, $\lambda_\theta = 0.1$ and $\lambda_\tau = 0.25$ such that in the hierarchical model of Van der Linden (2007, 2009), the correlation between τ_p and θ_p equals .37 and the standard deviation of τ_p equals 0.27 (see equations (4) and (5)). The residual intercept parameters, δ_{0i}, were equal to -1.0. The values above already provide all parameters for the GLM. Next, for the GLM-skew, λ_ζ was chosen to equal -0.025. This choice for λ_ζ given the choices for ν_i and λ_τ above

corresponds to a 'medium' effect on the skewness of the speed factor distribution according to Molenaar, Dolan, and Verhelst (2010) and Smits *et al.* (2015).[2] For the het-GLM, data are generated using $\delta_{1i} = -0.8$ for all i. This choice for δ_{1i} corresponds to a standardized value of δ_{1i} $\sigma_\tau = -0.8 \times 0.27 = -0.22$ which is a 'medium' effect according to Molenaar *et al.* (2010); standardizing δ_{1i} is necessary as its value depends on the scale of τ_p. Finally, for the full model, het-GLM-skew, data are generated using both $\lambda_\zeta = -0.025$ and $\delta_{1i} = -0.8$ for all i.

We conducted 100 replications for each true model. To each data set we fitted the four models discussed above. For all models we determined the AIC, BIC, saBIC, and the likelihood of the data given the model parameters. In addition, we conducted the multivariate normality test of Mardia (1970) to see how this test performs as compared to our modelling approach.

4.2. Results

4.2.1. Parameter recovery
Here, we focus on the recovery of the main parameters in the full model (M1: het-GLM-skew): the heteroscedasticity parameters, δ_{1i}, and the scale and shape parameters of the speed factor, λ_τ and λ_ζ. First, Table 1 contains the mean parameter estimates, mean standard errors, and the root mean squared error (*RMSE*) of λ_τ, λ_ζ, and δ_{1i} in the full model for the different true models. As can be seen, for all parameters, the results seem acceptable, that is, the parameters are acceptably recovered in all cases with the mean standard errors close to the *RMSE*, indicating no bias. That is, in the cases that the speed factor is skew normal (i.e., M1: het-GLM-skew and M3: GLM-skew), the mean estimates of λ_ζ are equal to their true values (-0.025) and in the cases where the speed factor is normal (i.e., M2: het-GLM and M4: GLM), the mean parameter estimates for λ_ζ are 0. A similar pattern holds for the recovery of δ_{1i}. That is, in the cases where the residuals are heteroscedastic (i.e., M1: het-GLM-skew and M2: het-GLM), the parameter estimates are close to their true value (-0.8), and in the cases that the residuals are homoscedastic (i.e., M3: GLM-skew and M4: GLM), the parameter estimates for δ_{1i} are close to 0. The results for λ_τ are also acceptable. Note that in the table, the results for δ_{1i} are aggregated over items. To present the results concerning these parameter estimates for each item separated we created box plots of the parameter estimates for δ_{1i} in the full model for the different true models in Figure 1.

4.2.2. Power and Type I error
We established the power to detect the different effects by determining the non-centrality parameter of the χ^2 distribution under the less restricted model using the procedure

---

[2] Smits *et al.* (2015) give algebraic expressions to transform ζ into λ_ζ for $v_i = 0$. As in our simulation study, v_i departs from 0, these expressions do not apply. However, we established that if

$$Z = 2 + 0.1X + 0.25Y - 0.025Y^2,$$

where $X \sim N(0,1)$ and $Y \sim N(0,1)$ (i.e., a regression model that corresponds to the design of the main simulation study discussed above where X represents θ_p and Y represents τ_p), the ζ estimate of Z over 100 replications has a mean of -2.06 ($SD = 0.26$). Note that Molenaar *et al.* (2010) defined $\zeta = 2.17$ as a medium effect size based on the standardized skewness coefficient of the skew normal distribution as proposed by Azzalini (1985). This value proposed by Molenaar *et al.* is thus close to the absolute value for ζ that is implied by our choice for λ_ζ in the design of the main simulation study (i.e., -0.025).

Table 1. True values, mean parameter estimates (*mEst*), mean standard errors (*mSE*), and root mean squared error (*RMSE*) of the λ_τ, λ_ζ, and δ_{1i} parameter estimates in the full model (M1: het-GLM-skew) for the different true models

True model	λ_τ				λ_ζ				δ_{1i}			
	True	mEst	mSE	RMSE	True	mEst	mSE	RMSE	True	mEst	mSE	RMSE
het-GLM-skew	0.250	0.248	0.011	0.013	−0.025	−0.025	0.007	0.007	−0.800	−0.811	0.299	0.300
het-GLM	0.250	0.251	0.011	0.011	0.000	0.000	0.007	0.008	−0.800	−0.796	0.297	0.299
GLM-skew	0.250	0.247	0.011	0.011	−0.025	−0.025	0.008	0.007	0.000	−0.007	0.300	0.296
GLM	0.250	0.250	0.010	0.010	0.000	0.000	0.007	0.007	0.000	−0.002	0.303	0.310

Note. The results of δ_{1i} are averaged over items. See Figure 1 for a graphical representation of the results for each item separately.
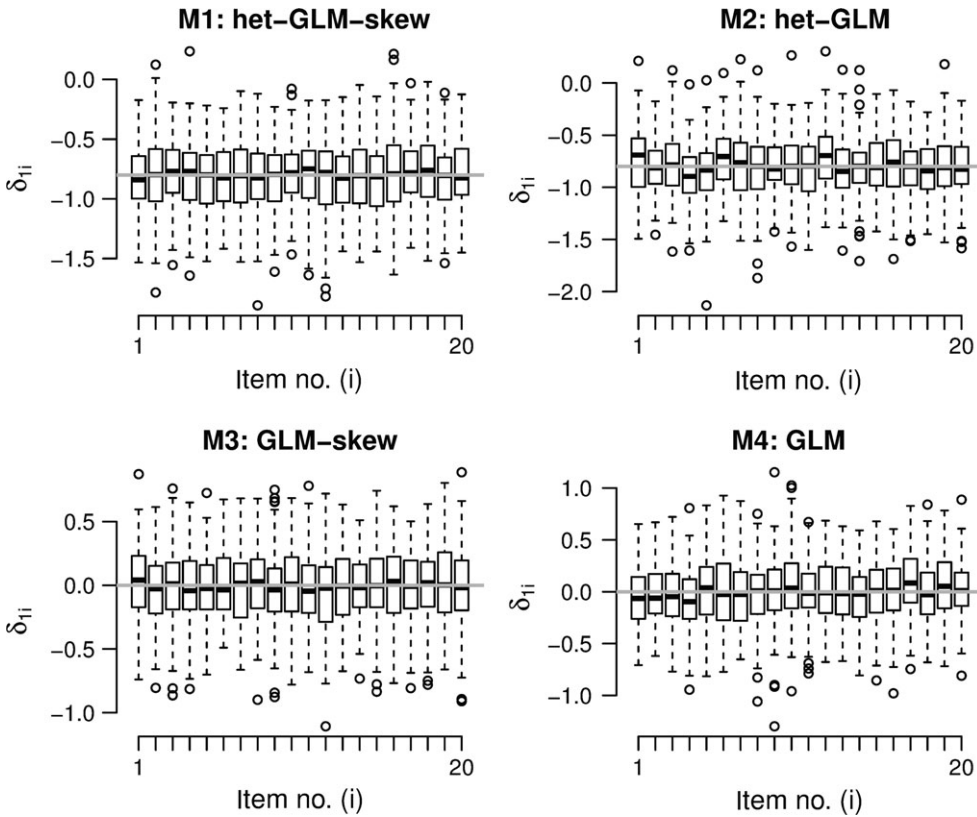
Figure 1. Box plots of the parameter estimates for δ_{1i} in the full model (M1: het-GLM-skew) across replications for the different true models. The grey solid line denotes the true values.

outlined by Satorra and Saris (1985; see also Saris & Satorra, 1993) with a .05 level of significance. For the full model (M1: het-GLM-skew), we separate between the power to detect heteroscedastic residuals in the presence of a skew normal speed factor (i.e., the power to reject $\delta_{1i} = 0$ for all i where λ_ζ is freely estimated) and the power to detect a skew normal speed factor in the presence of heteroscedasticity (i.e., the power to reject $\lambda_\zeta = 0$ where δ_{1i} is freely estimated for all i). For M2: het-GLM we only consider the power to detect heteroscedasticity (i.e., the power to reject $\delta_{1i} = 0$ for all i) and for M3: GLM-skew we only consider the power to detect skewness in the speed factor (i.e., the power to reject $\lambda_\zeta = 0$).

   The results concerning the power to detect the different effects are in Table 2. As can be seen, in the full model (M1: het-GLM-skew) both effects can be well separated. That is, if both effects are in the data (i.e., the true model is M1: het-GLM-skew) these are detected with a power close to 1.00, and if one of the effects is in the data, that effect is detected with an acceptable power, and the other effect (which is not in the data) is detected with a power close to the level of significance (indicating that the Type I error rate is not increased). For the model with heteroscedasticity only (M2: het-GLM), power and Type I error rate are also acceptable. That is, if there is heteroscedasticity in the data, it is well detected, and if the residuals are homoscedastic, power approaches the level of significance. For the model with a skew normal factor distribution only, the power to reject $\lambda_\zeta = 0$ is moderate (.68) if the data contain

Table 2. Power and non-centrality parameter (ncp) of the likelihood ratio test to reject $\delta_{1i} = 0$ and $\lambda_\zeta = 0$ in the different models at the .05 level of significance

| True model | M1: het-GLM-skew | | | | M2: het-GLM | | M3: GLM-skew | |
| | δ_{1i} | | λ_ζ | | | | | |
	power	ncp	power	ncp	power	ncp	power	ncp
M1: het-GLM-skew	1.00	141.27	0.95	12.95	1.00	134.15	0.68	5.83
M2: het-GLM	1.00	142.36	0.08	0.23	1.00	143.57	0.22	1.44
M3: GLM-skew	0.05	0.00*	0.93	11.98	0.05	0.00*	0.94	12.11
M4: GLM	0.06	0.59	0.05	0.00*	0.06	0.59	0.05	0.00*

Note. *These ncp values are fixed to 0 as their estimate was slightly negative (which can happen due to sampling fluctuations).

heteroscedasticity residuals as well (i.e., the true model is M1: het-GLM-skew). In addition, the Type I error rate seems increased. That is, if the data contain heteroscedastic residuals but no skewness in the speed factor distribution (i.e., the true model is M2: het-GLM), the power to reject $\lambda_\zeta = 0$ is still .22.

With respect to the fit indices, we focused on the hit rates and false positive rates of the different models for the different effects in the data. That is, the hit rate is defined as the proportion of replications in which the true model is correctly identified as the best model according to the AIC, BIC and saBIC, and the false positive rate is defined as the proportion of replications in which the wrong model is identified as the best model. See Table 3 for the results. As can be seen, all fit indices perform acceptably in terms of the hit rates. In addition, the effects are statistically well separable, meaning that heteroscedastic residuals are seldom detected as skew normal factors or vice versa (i.e., the off-diagonal proportions are close to zero for het-GLM and GLM-skew). This emphasizes the need to consider both effects: they pick up different aspects in the data.

For each replication in the simulation study, we also conducted the test by Mardia (1970) on multivariate normality. Results indicate that non-normality was rejected in 98, 99, 5 and 2% of the replications if the true model was M1: het-GLM-skew, M2: het-GLM, M3: GLM-skew and M4: GLM, respectively. Thus, normality was rejected if the data contained both effects, and the effect of heteroscedasticity only, but normality was not rejected when the data contained a skewed speed factor.

5. Illustration

5.1. Data

The data comprise the responses and response times of 668 Dutch high school students on 23 items from the 'path finder' test. This test resembles the embedded figures test (Witkin, 1950). In the test, respondents have to decide which of five simple figures occur within a more complex target figure. Two of the respondents were omitted from the analyses because they showed suspiciously small response times (1 s on most of the items). Following the simulation study, we focus on the log-transformed response times. A time deadline of 40 s was imposed on all items. As a result, the observed response times showed ceiling effects which caused the log response time distribution to be non-normal. To illustrate how this source of non-normality can be detected and accounted for using the present approach we fit the different models as studied in the simulation study to the data.

Table 3. Hit rates and false positive rates for the AIC, BIC and saBIC for the estimated models across the different true models

	True model	Estimated models			
		het-GLM-skew	het-GLM	GLM-skew	GLM
AIC	het-GLM-skew	**.99**	.01	0	0
	het-GLM	.23	**.77**	0	0
	GLM-skew	0	0	**.97**	.03
	GLM	0	.01	.12	**.87**
BIC	het-GLM-skew	**.73**	.14	.05	.08
	het-GLM	.02	**.93**	0	.05
	GLM-skew	0	0	**.82**	.18
	GLM	0	0	.01	**.99**
saBIC	het-GLM-skew	**.95**	.05	0	0
	het-GLM	.13	**.87**	0	0
	GLM-skew	0	0	**.94**	.06
	GLM	0	0	.07	**.93**

Note. Hit rates are in boldface (i.e., the proportion of replications in which the true model is correctly identified as the best model).

Next, we show how estimates for θ_p and τ_p differ between the present non-normal models and the more traditional normal baseline model.

5.2. Results

Table 4 contains the results concerning the fit of the different models. As can be seen, the likelihood ratio test between each of the models and the full model (M1: het-GLM-skew) is only significant for the model with heteroscedasticity only, M2: het-GLM, $\chi^2(1) = .74$, $p = .39$, indicating that dropping the skewness parameter, λ_ζ, from the full model does not deteriorate model fit. As both the model with skewness only (M3: GLM-skew) and the model without effects (M4: GLM) are associated with a significant likelihood ratio test, we choose M2: het-GLM as the best-fitting model. This conclusion is supported by the AIC, BIC, and saBIC. See Table 5 for the item parameter estimates within this model. As can be seen, $\delta_{1i} < 0$ for all i, indicating that the residual variance is smaller for subjects with lower levels of τ_p. This is likely due to the item deadline which results in less variance at the upper range of the log response time distribution.

Table 4. Likelihood ratio test with M1: het-GLM-skew and fit indices for the models considered in the illustration

	ℓ	LRT	df	AIC	BIC	saBIC
M1: het-GLM-skew	−20,914.49	–	–	42,065	42,596	42,221
M2: het-GLM	−20,915.23	0.74	1	**42,064**	**42,591**	**42,220**
M3: GLM-skew	−21,740.77	826.28	23	43,672	44,099	43,798
M4: GLM	−21,785.36	870.87	24	43,759	44,182	43,883

*Note.* For the fit indices, best values are in bold.

Table 5. Parameters estimates (est) and standard errors (*SE*) for the parameters in M2: het-GLM

| α_i		β_i		ν_i		δ_{0i}		δ_{1i}	
est	*SE*	est	*SE*	est	*SE*	est	*SE*	est	*SE*
0.65	0.10	−0.88	0.09	2.44	0.04	−0.44	0.05	−0.45	0.28
0.61	0.11	−0.02	0.08	2.51	0.03	−0.79	0.05	−0.23	0.33
0.80	0.11	−1.04	0.10	2.60	0.03	−0.73	0.05	−0.83	0.29
0.25	0.09	−1.06	0.09	2.71	0.03	−1.01	0.06	−0.65	0.38
0.36	0.12	1.04	0.09	3.04	0.03	−2.16	0.10	−4.25	0.58
0.54	0.13	0.96	0.09	2.75	0.04	−1.40	0.09	−2.73	0.40
0.15	0.09	0.49	0.08	3.02	0.03	−1.91	0.09	−2.56	0.60
0.40	0.11	0.71	0.09	2.98	0.03	−1.98	0.08	−3.70	0.84
0.36	0.12	1.11	0.09	2.95	0.03	−1.81	0.12	−2.79	1.00
0.39	0.09	0.25	0.08	2.93	0.03	−1.81	0.09	−3.31	0.96
0.20	0.09	0.60	0.08	2.92	0.03	−1.84	0.08	−2.52	0.51
0.22	0.12	1.21	0.09	2.80	0.03	−1.39	0.09	−3.61	0.55
0.53	0.11	0.19	0.08	2.73	0.03	−1.34	0.08	−1.50	0.38
0.48	0.11	0.39	0.08	2.85	0.04	−1.58	0.11	−2.54	0.57
0.53	0.11	0.70	0.09	2.88	0.04	−1.62	0.12	−3.34	0.93
0.52	0.10	0.16	0.08	2.94	0.03	−1.47	0.09	−1.81	0.59
0.40	0.12	0.90	0.09	2.93	0.02	−1.81	0.09	−3.28	1.20
0.36	0.11	0.62	0.08	2.93	0.03	−2.02	0.11	−4.37	0.85
0.66	0.17	1.05	0.10	2.82	0.03	−1.48	0.08	−3.20	0.63
0.29	0.10	0.81	0.09	2.79	0.03	−1.39	0.08	−3.22	0.52

To see how the effect of heteroscedasticity affects the results of the normal baseline model, we plot the estimates of θ_p and τ_p in the final model, M2: het-GLM, against the estimates of respectively θ_p and τ_p in the baseline model, M4: GLM (see Figures 2 and 3). As can be seen, the estimates are highly comparable in the mid-range of θ_p and τ_p. However, in the tails, the estimates diverge. This is because the non-normality in the log response time distribution is most evident in the tails of the distribution (in this case, a thicker lower tail and a thinner upper tail). As M2: het-GLM fits the data best according to the fit indices, the θ_p and τ_p estimates from this model can be better trusted than those from the baseline model. Especially in an applied setting where the interest is in identifying high- or low-scoring respondents (e.g., personnel selection or admission for extra training for weak students), the θ_p estimates from the non-normal approach might be the preferred basis for inferences.

6. Discussion

In this paper, we have presented a method to test for departures from a normal distribution for the transformed response times within the GLM framework for responses and response times (e.g., Van der Linden, 2007). Most importantly, we distinguished between non-normality due to a non-normal speed factor and non-normality due to heteroscedastic residual variances. Our simulation study showed that this distinction may be valuable as the power to detect violations of normality in the transformed response times due to a non-normal speed factor is small for the more traditional marginal test of Mardia (1970) on multivariate normality. Molenaar *et al.* (2010) found a similar result within the one-factor model.
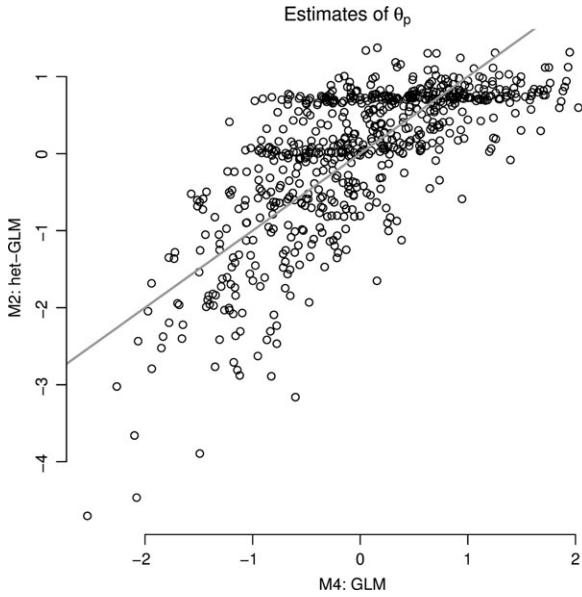
Figure 2. Plot of estimates of θ_p in M2: het-GLM (y-axis) against M4: GLM (x-axis). The solid grey line denotes a one-to-one correspondence.
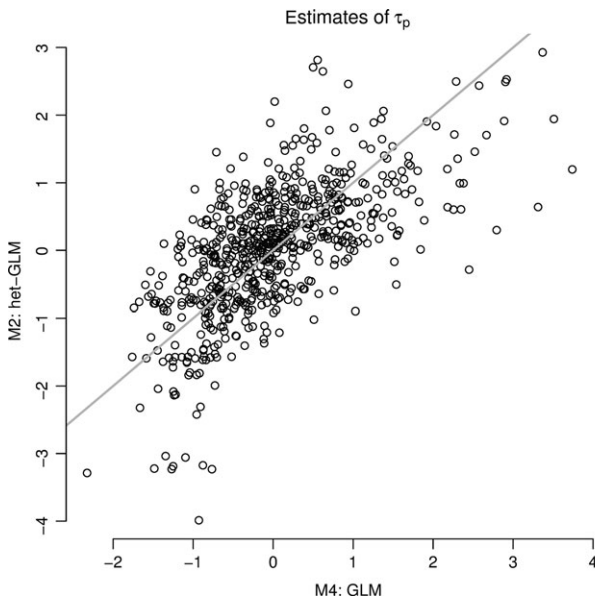
Figure 3. Plot of estimates of τ_p in M2: het-GLM (y-axis) against M4: GLM (x-axis). The solid grey line denotes a one-to-one correspondence.

Our main motivation to adopt the present modelling approach was, first, to have a statistical tool available to study whether results from the traditional normal approaches are affected by non-normality in the transformed response times; and second, to test for

qualitative differences in response behaviour, including differences in solution strategies, guessing, etc. With respect to the latter motivation, differences in response behaviour may be detected because they will produce a mixture of two or more qualitatively different classes of respondents. In the present approach, the presence of different classes in a given data set can be detected in the speed factor distribution (indicating between-subject differences) or in the residuals (indicating within-subject differences). However, the exact number of classes cannot be determined. That is, violations of normality may indicate qualitative differences in the response behaviour of the respondents, but the results give no indication about the nature or number of classes.

If qualitative differences in the response behaviour of the respondents result in a multi-modal mixture, the present approach can still be used as a statistical test for the presence of such differences, but it cannot be used to actually model the bimodality in the distribution. In addition, if the mixture is perfectly multi-modal (i.e., the mixture distribution is symmetrical), the present modelling approach will not be suitable as the observed data distribution will not be skewed. However, in less perfect cases (e.g., one mixture component is smaller than the other or one mixture component has smaller variance), the multi-modality that arises will be characterized by skewness and our approach can be used to test for this effect, given that the effect size is large enough.

Acknowledgements

References

Aberson, C. L., Shoemaker, C., & Tomolillo, C. (2004). Implicit bias and contact: The role of interethnic friendships. *Journal of Social Psychology*, *144*, 335–347. doi:10.3200/SOCP.144.3.335-347

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705

Andersen, E., & Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika*, *42*, 357–374. doi:10.1007/BF02293656

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, *62*, 245–253. doi:10.1111/j.1541-0420.2005.00440.x

Azevedo, C. L. N., Bolfarine, H., & Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Computational Statistics and Data Analysis*, *55*, 353–365. doi:10.1016/j.csda.2010.05.003

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178. www.jstor.org/stable/4615982

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, *46*, 199–208. doi:10.6092/issn.1973-2201/711

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459. doi: 10.1007/BF02293801

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.

Brunner, E., Dette, H., & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, *92*, 1494–1502. doi:10.1080/01621459.1997.10473671

Ferrando, P. J., & Lorenzo-Seva, U. (2007a). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement*, *31*, 525–543. doi:10.1177/0146621606295197

Ferrando, P. J., & Lorenzo-Seva, U. (2007b). A measurement model for Likert responses that incorporates response time. *Multivariate Behavioral Research*, *42*, 675–706. doi:10.1080/00273170701710247

Furneaux, W. D. (1961). Intellectual abilities and problem solving behavior. In H. J. Eysenck (Ed.), *The handbook of abnormal psychology* (pp. 212–237). London, UK: Pitman.

Glas, C. A., & van der Linden, W. J. (2010). Marginal likelihood inference for a model for item responses and response times. *British Journal of Mathematical and Statistical Psychology*, *63*, 603–626. doi:10.1348/000711009X481360

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica: Journal of the Econometric Society*, *44*, 461–465. doi:10.2307/1913974

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

Holden, R. R., & Kroner, D. G. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment*, *4*, 170–173. doi:10.1037/1040-3590.4.2.170

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*, 255–259. doi:10.1016/0165-1765(80)90024-5

Klein Entink, R. H., Linden, W. J., & Fox, J. P. (2009). A Box-Cox normal model for response times. *British Journal of Mathematical and Statistical Psychology*, *62*, 621–640. doi:10.1348/000711008X374126

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. doi:10.1007/BF02296338

Kuiper, N. A. (1981). Convergent evidence for the self as a prototype: The 'inverted-U RT effect' for self and other judgments. *Personality and Social Psychology Bulletin*, 7, 438–443. doi:10.1177/014616728173012

Lewin-Koh, S., & Amemiya, Y. (2003). Heteroscedastic factor analysis. *Biometrika*, *90*, 85–97. doi:10.1093/biomet/90.1.85

Loeys, T., Legrand, C., Schettino, A., & Pourtois, G. (2014). Semi-parametric proportional hazards models with crossed random effects for psychometric response times. *British Journal of Mathematical and Statistical Psychology*, *67*, 304–327. doi:10.1111/bmsp.12020

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician*, *54*, 217–224.

Luce, R. D. (1986). *Response times*. New York, NY: Oxford University Press.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530. doi:10.1093/biomet/57.3.519

Marianti, S., Fox, J. P., Avetisyan, M., Veldkamp, B. P., & Tijmstra, J. (2014). Testing for aberrant behavior in response time modeling. *Journal of Educational and Behavioral Statistics*, *39*, 426–451. doi:10.3102/1076998614559412

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, *27*, 392–415. doi:10.1007/BF02289646

McLeod, L., Lewis, C., & Thissen, D. (2003). A Bayesian method for the detection of item preknowledge in computerized adaptive testing. *Applied Psychological Measurement*, *27*(2), 121–137. doi:10.1177/0146621602250534

Meijer, E., & Mooijaart, A. (1996). Factor analysis with heteroscedastic errors. *British Journal of Mathematical and Statistical Psychology*, *49*, 189–202. doi:10.1111/j.2044-8317.1996.tb01082.x

Meyer, J. P. (2010). A mixture Rasch model with item response time components. *Applied Psychological Measurement*, *34*, 521–538. doi:10.1177/0146621609355451

Molenaar, D., Dolan, C. V., & de Boeck, P. (2012). The heteroscedastic graded response model with a skewed latent trait: Testing statistical and substantive hypotheses related to skewed item category functions. *Psychometrika*, *77*, 455–478. doi:10.1007/s11336-012-9273-5

Molenaar, D., Dolan, C. V., & Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology*, *63*, 293–317. doi:10.1348/000711009X456935

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. J. (2015a). A bivariate generalized linear item response theory modeling framework to the analysis of responses and response times. *Multivariate Behavioral Research*, *50*, 56–74. doi:10.1080/00273171.2014.962684

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. J. (2015b). A generalized linear factor model approach to the hierarchical framework for responses and response times. *British Journal of Mathematical and Statistical Psychology*, *68*, 197–219. doi:10.1111/bmsp.12042

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

Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology*, *70*, 305–311. doi:10.1111/j.2044-8295.1979.tb01687.x

Ranger, J. (2013). Modeling responses and response times in personality tests with rating scales. *Psychological Test and Assessment Modeling*, *55*, 361–382.

Ranger, J., & Kuhn, J. T. (2012). A flexible latent trait model for response times in tests. *Psychometrika*, *77*, 31–47. doi:10.1007/s11336-011-9231-7

Ranger, J., & Kuhn, J. T. (2013). Analyzing response times in tests with rank correlation approaches. *Journal of Educational and Behavioral Statistics*, *38*(1), 61–80. doi:10.3102/1076998611431086.

Ranger, J., & Ortner, T. (2012a). The case of dependency of responses and response times: A modeling approach based on standard latent trait models. *Psychological Test and Assessment Modeling*, *54*, 128–148.

Ranger, J., & Ortner, T. (2012b). A latent trait model for response times on tests employing the proportional hazards model. *British Journal of Mathematical and Statistical Psychology*, *65*, 334–349. doi:10.1111/j.2044-8317.2011.02032.x

Ranger, J., & Ortner, T. M. (2013). Response time modeling based on the proportional hazards model. *Multivariate Behavioral Research*, *48*, 503–533. doi:10.1080/00273171.2013.796280

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. doi:10.1007/BF02294150

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 Behavioral Research*, *41*, 427–443. doi:10.1207/s15327906mbr4104_1

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343. doi:10.1007/BF02294360

Shapiro, S. S., & Wilks, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611. www.jstor.org/stable/2333709

Smits, I. A., Timmerman, M. E., & Stegeman, A. (2015). Modelling non-normal data: The relationship between the skew-normal factor model and the quadratic factor model. *British Journal of Mathematical and Statistical Psychology*, *69*, 105–121. doi:10.1111/bmsp.12062

Stevens, J. (2009). *Applied multivariate statistics for the social sciences*. New York, NY: Routledge.

Thissen, D. (1983). Timed testing: An approach using item response testing. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait theory and computerized adaptive testing* (pp. 179–203). New York, NY: Academic Press.

Van den Oord, E. J. C. G. (2005). Estimating Johnson curve population distributions in MULTILOG. *Applied Psychological Measurement*, *29*, 45–64. doi:10.1177/0146621604269791

Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 287–308. doi:10.1007/s11336-006-1478-z

Van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *46*, 247–272. doi:10.1111/j.1745-3984.2009.00080.x

Van der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in multistage testing. *Journal of Educational Measurement*, *44*(2), 117–130. doi:10.1111/j.1745-3984.2007.00030.x

Van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, *73*, 365–384. doi:10.1007/s11336-007-9046-8

Van der Linden, W. J., Scrams, D. J., & Schnipke, D. L. (1999). Using response-time constraints to control for differential speededness in computerized adaptive testing. *Applied Psychological Measurement*, *23*, 195–210. doi:10.1177/01466219922031329

Van der Linden, W. J., & van Krimpen-Stoop, E. M. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, *68*, 251–265. doi:10.1007/BF02294800

Van der Maas, H. L., & Jansen, B. R. (2003). What response times tell of children's behavior on the balance scale task. *Journal of Experimental Child Psychology*, *85*, 141–177. doi:10.1016/S0022-0965(03)00058-4

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, *58*, 220–233. doi:10.1046/j.0039-0402.2003.00257.x

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, UK: Cambridge University Press.

Wang, C., Chang, H. H., & Douglas, J. A. (2013). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology*, *66*, 144–168. doi:10.1111/j.2044-8317.2012.02045.x

Wang, C., Fan, Z., Chang, H. H., & Douglas, J. A. (2013). A semiparametric model for jointly analyzing response times and accuracy in computerized testing. *Journal of Educational and Behavioral Statistics*, *38*, 381–417. doi:10.3102/1076998612461831

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*, 456–477. doi:10.1111/bmsp.12054

Witkin, H. A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality*, *19*, 1–15. doi:10.1111/j.1467-6494.1950.tb01084.x

Woods, C. M. (2007). Ramsay-curve IRT for Likert type data. *Applied Psychological Measurement*, *31*, 195–212. doi:10.1177/0146621606291567

Supporting Information

The following supporting information may be found in the online edition of the article:

Data S1. Application files.

Data S2. Building Mplus scripts.

Data S3. Simulation files.