Nonlinear indicator-level moderation in latent variable models

May 31, 2018

Abstract

Linear, nonlinear, and nonparametric moderated latent variable models have been developed to investigate possible interaction effects between a latent variable and an external continuous moderator on the observed indicators in the latent variable model. Most moderation models have focused on moderators that vary across persons but not across the indicators (e.g., moderators like age and SES). However, in many applications, the values of the moderator may vary both across persons and across indicators (e.g., moderators like response times and confidence ratings). Indicator-level moderation models are available for categorical moderators and linear interaction effects. However these approaches require, respectively categorization of the continuous moderator and the assumption of linearity of the interaction effect. In this paper, parametric nonlinear and nonparametric indicator-level moderation methods are developed. In a simulation study we demonstrate the viability of these

1

methods. In addition, the methods are applied to a real dataset pertaining to arithmetic ability.

1 Introduction

Traditional latent variable models have been used to operationalize a psychological construct in terms of its observed indicators. The general idea is that the indicators are regressed on a latent variable, representing the construct, resulting in indicator-specific intercept and slope parameters (and in the case of continuous indicators the indicator-specific residual variance parameters). Well-established latent variable models exist for dichotomous indicators (e.g., the 2-parameter normal ogive model; 2PNOM, Lord & Novick, 1968), ordinal indicators (e.g., the graded response model, Samejima, 1969), and continuous indicators (e.g., the linear factor model, Spearman, 1904).

Moderated latent variable models

With these latent variable models in place, there ~~us~~ is growing interest in testing for possible interaction effects between the latent variable and an external moderation variable (e.g., gender or age) on the observed indicators. We see two reasons for the development of these so-called moderated latent variable models:—. First, tests on moderation of the parameters in a latent variable model, like the 2PNOM or the linear factor model, constitute the primary idea of tests on measurement invariance (Meredith, 1964, 1993; Muthén, 1989). Measurement invariance (or, the absence of differential item functioning, Mellenbergh, 1989) refers to the prerequisite that the

2

parameters from a latent variable model should not be moderated by a background variable, as this will hamper ~~a~~ the comparison of subjects' observed indicator scores in terms of the underlying latent variable. Although the initial developments of the statistical toolkit for establishing measurement invariance ~~have not explicitly been~~ were not explicitly conducted in a moderated latent variable modeling framework, recently, this correspondence ~~was~~ has been made explicit by Bauer (2017) and Curran et al. (2014).

Another reason for the development of moderated latent variable models is a substantive one. That is, theories from various applied fields predict interactions between observed and latent variables, a notion that can statistically be properly tested using a moderated latent variable model. For instance, in the field of behavior genetics, it has been found using moderated latent variable models that the heritability of intelligence, which is a latent variable, is moderated by socioeconomic status (SES) in such a way that intelligence is more heritable for subjects of higher SES (e.g., Turkheimer, Haley, Waldron, d'Onofrio, & Gottesman, 2003). In addition, in the field of intelligence research, age differentiation (Garrett, 1946) refers to the prediction that intelligence is a weaker source of individual difference for older ages. This prediction has been tested by Facon (2006) using categorized age~~as moderator~~, and by Tucker-Drob (2009) using continuous age as a moderator. Another example is from personality research ~~were~~ where it has been hypothesized that personality is moderated by IQ with less personality variance for higher levels of IQ. In Murray, Booth, and Molenaar (2016), this hypothesis was tested using a moderated latent variable model, but no overall evidence for the personality-IQ interaction was found.

Traditional moderation approaches

If we accept the importance of moderated latent variable models in tests on measurement invariance and substantive hypotheses, question arises as to which moderation approaches are readily available. ~~See Table ?? for an overview.~~ In the case of discrete moderators like gender, tests on the interaction effect between the latent variable and the moderator are straightforward using multi-group latent variable models for continuous data (e.g., Jöreskog, 1971), ordinal data (e.g., Lee, Poon, & Bentler, 1989), or dichotomous data (e.g., Muthen & Christoffersson, 1981). In these models, differences in the intercept parameters between groups reflect a main effect of the moderator on the observed indicators, and ~~a difference~~ differences in the slope parameters between groups reflect an interaction between the latent variable and the moderator on the observed indicators. In the context of dichotomous data, various item response theory based methods have been developed for testing whether the model parameters differ between groups (e.g., Lord, 1980; Mellenbergh, 1989; Thissen, Steinberg, & Gerrard, 1986) which is usually referred to as differential item functioning (DIF) or item bias ~~(see Millsap, 2011, for an overview of the existing approaches to test for DIF)~~(see Millsap, 2

In the case of continuous moderators, an intuitive approach is to categorize the continuous moderator variable and test ~~for~~ the differences in the intercepts and slope parameters using the multi-group methods~~above~~. However, several authors have argued against categorization of continuous variables (see e.g., Cohen, 1983; MacCallum, Zhang, Preacher, & Rucker, 2002; McClelland, Lynch, Irwin, Spiller, & Fitzsimons, 2015) as it lowers the information concerning individual differences ~~affecting~~ and therefore affects the power to detect a possible effect of that variable.

4

In addition, cut-off points are arbitrary, which may complicate the comparison of results across studies (Royston, Altman, & Sauerbrei, 2006). Therefore, effort has been devoted to develop models for continuous moderators which avoid the necessity of categorization. Latent variable models have been proposed in which the intercept and slope parameters are a linear function of the moderator, resulting in a baseline parameter and a moderation parameter for both the intercept and the slope ~~parameter~~ parameters (Purcell, 2002; Rabe-Hesketh, Skrondal, & Pickles, 2004; Mehta & Neale, 2005; Neale, Aggen, Maes, Kubarych, & Schmitt, 2006; Bauer & Hussong, 2009; Molenaar, Dolan, Wicherts, & van der Maas, 2010). The moderation parameter of the intercept reflects the main effect of the moderator, and the moderation parameter of the slope parameter reflects the linear interaction effect.

In the ~~linear~~ nonlinear moderation model, the assumption of a linear dependence of the intercept and slope parameters on the moderator variable can be relaxed. That is, adding higher-order powers of the moderator as additional moderator variables enables tests on quadratic or cubic interactions (see e.g., Purcell, 2002; Tucker-Drob, 2009). However, one is still tied to the parametric form of the interaction. To explore the relation between the intercept and slope parameters on the one hand, and the moderator on the other, research has focused on nonparametric approaches. Specifically, Hildebrandt, Wilhelm, and Robitzsch (2009); Hülür, Wilhelm, and Robitzsch (2011) and Hildebrandt, Lüdtke, Robitzsch, Sommer, and Wilhelm (2016) proposed a latent variable model in which the dependence of the slope and intercept parameter on the moderator variable is estimated using methodology from nonparametric regression analysis (Fox, 2015; Wu & Zhang, 2006). In an application of these meth-

ods, Briley, Harden, Bates, and Tucker-Drob (2015) found the heritability by SES interaction discussed above to be characterized by substantial nonlinear shifts, while the interaction was commonly assumed to be strictly linear before that.

Indicator-level approaches

Interestingly, all moderation models above have focused on moderators that vary across persons but not across the indicators (e.g., moderators like age and SES). However, in (potential) applications, the values of the moderator may vary both across persons and across indicators. For instance, in intelligence research, Partchev and de Boeck (2012) and DiTrapani, Jeon, De Boeck, and Partchev (2016) hypothesized that the item-specific response times of a cognitive ability test interact with the latent cognitive ability variable reflecting that respondents alternate between controlled processes (the slower responses) and automated processes (the faster responses). As respondents may respond relatively fast on one item (indicating an automated process) and relatively slow on the next (indicating controlled process), a person specific moderator (e.g., the mean response time of the full test) will not suffice in the analysis. That is, the response times constitute indicator-specific moderators that account for indicator-specific moderation effects. Other existing examples of indicator-level moderators include measures of answer changes (Jeon, De Boeck, & van der Linden, 2017) and confidence ratings (Gvozdenko, 2010). In addition, indicator-level moderators can be constituted by, for instance, verbally reported response processes, number of actions in interactive items, number of item clicks, number of eye fixations on the areas of interest, inspection times, response changes, certainty scores,

or physiological measures.

The development of methods to test for interactions between such indicator-level moderators and the latent variable has recently started to evolve across similar lines as described above for the traditional moderation models. ~~That is, approaches~~ Approaches have been proposed that require categorization of the indicator-level moderators (Partchev & de Boeck, 2012; DiTrapani et al., 2016) and models have been proposed by specifying linear functions between the intercept and slope parameter and the indicator-level moderator (Bolsinova, Tijmstra, & Molenaar, 2017; Goldhammer, Steinwascher, Kroehne, & Naumann, 2017; Bolsinova, de Boeck, & Tijmstra, 2017). ~~That is, these~~ These existing indicator-level approaches focus only on categorical moderators, or continuous linear moderators. However, as with the traditional moderation models, the assumption of linearity of indicator-level moderation models might be violated in practice. In such cases, using linear models may result in invalid conclusions about the relationship between the parameters of the latent variable model and the indicator-specific moderator. For instance, one might conclude that the intercept increases with the values of the moderator (e.g., that slower responses on an intelligence test are more often correct), while it might be that it increases only up to some value of the moderator and decreases after that value, or that the increase is not linear. However, parametric nonlinear and nonparametric approaches are not yet developed for indicator-specific moderation ~~(Table ??)~~. Therefore, in this paper we propose modeling the relationship between the indicator-specific moderator and the parameters of the latent variable model in a more flexible way. ~~That is, we propose a parametric nonlinear and nonparametric approach for~~

7

indicator-level moderation.

Developing a parametric nonlinear approach and a nonparametric indicator-level approach

The development of parametric indicator-level nonlinear models is relatively straightforward as the existing parametric indicator-level linear moderation approach can be extended to include higher-order interactions. However, as these interactions are indicator-specific, it is not obvious that the resulting model is identified and performs satisfactorily in recovering the true relation between the moderator and the indicator data. In addition, development of the nonparametric indicator-level moderation approach is not straightforward, as the traditional nonparametric moderation approach cannot simply be extended to include one moderator for each indicator. That is, the The nonparametric moderation is based on a binning procedure which is feasible because there is only one moderator. In the case of indicator-specific moderators however, such a binning procedure will result in a grid where the dimensions of the grid grow exponentially with the number of indicators. Such a procedure is infeasible in practice. Taken together, we therefore think that studying the parametric nonlinear extension and a feasible nonparametric approach to indicator-level moderation is a worthwhile endeavor. The outline of this paper is as follows: . First, the existing moderated latent variable models are formally presented. Next, the parametric nonlinear and nonparametric approach for indicator-level moderation will be derived and estimation for these new moderation models will be discussed. Each of these approaches will be applied to a dataset including both the responses and the

response times to an arithmetic test. Then, these results will be used in a simulation study to demonstrate the viability of the indicator-level moderation models. We end with a general discussion.

2 Moderated Latent Variable Models: Existing approaches

In the traditional (non-moderated) generalized linear latent variable model, the conditional expectation of the i-th observed indicator, denoted by Y_i, given the latent variable, denoted by η, is specified via a link function $g(\cdot)$ and a linear function:

$$g[E(Y_i \,|\, \eta)] = \lambda_i \eta + \nu_i, \tag{1}$$

where λ_i and ν_i are the indicator-specific slope (also referred to as the factor loading) and intercept. The slope parameter indicates the strength of the relationship between the observed indicator and the latent variable. The intercept parameter determines the conditional expectation of the indicator when the latent variable is equal to zero, which is usually chosen to be equal to the expected value of the latent variable in the population. Various latent variable models are special cases of the general model above, depending on the choice for the link function, $g(.)$, and the shape of the distribution for the residuals (see Mellenbergh, 1994). For instance, if the link function is taken to be the identity link, the model above is equal to the linear factor model for continuous indicators; in the case of a probit link, the two-parameter

normal-ogive model for binary indicators arises, and in the case of a cumulative probit link, the model simplifies to a graded response model for ordinal indicators. ~~Note, that in the case of~~ For ordinal indicators with C categories, there are $C - 1$ intercept parameters for each indicator.

Let $\mathbf{y}$ denote an $N \times K$ data matrix with the observed ~~indicators~~ data of N persons ~~to~~ on K indicators, and ~~by~~ y_{pi} the value of the i-th indicator from the p-th person (i.e., a realisation of the random variable Y_i for person p). In latent variable models, conditional independence of the observed indicators given the latent variables is typically assumed. The marginal likelihood of the data matrix $\mathbf{y}$ can be computed as follows

$$f(\mathbf{y} \mid \boldsymbol{\lambda}, \boldsymbol{\nu}) = \prod_{p=1}^{N} \int \prod_{i=1}^{K} f(y_{pi}; \lambda_i, \nu_i, \eta) u(\eta) d\eta, \tag{2}$$

where $f(y_{pi}; \lambda_i, \nu_i, \eta)$ is the conditional density of the observation y_{pi} given the latent variable and $u(\eta)$ is the population distribution of the latent variable, which is often assumed to be standard normal.[1] In the case of the linear factor model the conditional density is assumed to be normal:

$$f(y_{pi}; \lambda_i, \nu_i, \eta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ -\frac{(y_{pi} - (\lambda_i \eta + \nu_i))^2}{2\sigma_i^2} \right\}, \tag{3}$$

where σ_i^2 is the additional indicator-specific parameter which determines the conditional variance of the indicator given the latent variable (i.e., the residual variance). In the case of the graded response model for ordinal indicators, the conditional den-

---

[1]Note that the mean and the variance of the latent variable distribution are constrained for identification purposes. Alternatively, the parameters of Y_i can be constrained for some arbitrary i, that is, $\lambda_i = c_1, \nu_i = c_2$, where $c_1 > 0$ and c_2 are arbitrary values.

sity is multinomial:

$$f(y_{pi}; \lambda_i, \boldsymbol{\nu}_i, \eta) = \prod_{c=0}^{C} \underline{(} \left\{ \Phi(\lambda_i\eta + \nu_{ic}) - \Phi(\lambda_i\eta + \nu_{i(c-1)}) \underline{)} \right\}^{\mathcal{I}(y_{pi}=c)} \tag{4}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, the category-specific intercept parameters are ordered within each indicator, $\nu_{i0} = \infty$ and $\nu_{iC} = -\infty$. In the case of the 2PNOM for binary indicators, the conditional density is Bernoulli:

$$f(y_{pi}; \lambda_i, \nu_i, \eta) = \Phi(\lambda_i\eta + \nu_i)^{y_{pi}} (1 - \Phi(\lambda_i\eta + \nu_i))^{1-y_{pi}}. \tag{5}$$

Moderation can be introduced in the generalized linear latent variable model above by making the slope and intercept parameters ~~above~~ a function of the moderator, Z, that is (see e.g., Hildebrandt et al., 2016):

$$g[E(Y_i \,|\, \eta, Z)] = f_{\lambda_i}(Z)\eta + f_{\nu_i}(Z). \tag{6}$$

with

$$VAR(Y_i \,|\, Z) = f_{\ln[\sigma_i^2]}(Z) \tag{7}$$

in the case of continuous Y_i. ~~In the above,~~ $f_{\lambda_i}(.)$, $f_{\nu_i}(.)$, and $f_{\ln[\sigma_i]}(.)$ model the functional relationship between the <u>moderator and the</u> slope parameters, the intercept parameters, and the log-residual variance~~on the one hand and the moderator on the other~~, <u>respectively</u>. Note that we model the log-residual variance to ensure that the residual variance itself is strictly positive.

Equation 6 is a general form of a traditional moderation model, ~~that is a model~~

11

in which the value of the moderator varies across persons and the same moderator applies to all indicator variables. In indicator-level moderation models, it is assumed that there are as many moderators as there are indicator variables (K). Congruently to the person-level approach above, in the indicator-level moderation model the slope and the intercept parameters depend not on the value of the moderator Z, common for all the indicators, but on the value of the moderator which is now indicator-specific, denoted by Z_i. To obtain the general form of an indicator-level moderation model, one needs to substitute Z with Z_i in Equation 6. ~~Note, that~~ The indicator-level moderation models assume that the parameters of each indicator only depend on the corresponding moderator, ~~that is~~ and independence is assumed between Z_i and Y_k for all $i \neq k$.

2.1 Model interpretation

In the general moderation model in Equation 6, the main parameters of the latent variable model are a function of a moderator. For the intercepts, this effect reflects the main effect of the moderator on the observed indicators. In most applications, this effect is not of primary interest. For instance, in the intelligence by age interaction example ~~discussed above~~ (i.e., age differentiation; Tucker-Drob, 2009), the intercepts in the latent variable model are commonly found to be moderated by age. This simply reflects that there is a correlation between intelligence and age, which is well known. However, if age moderates the slope parameters of the latent variable model in such a way that the slopes are smaller for higher ages, this reflects that the latent variable 'intelligence' manifests itself differently across age with smaller

intelligence variance for higher ages. Thus, ~~commonly,~~ the key interest in moderation models is to establish moderation of the slope parameters to study the interaction between the latent variable and the moderator. Moderation of the intercepts, although not of key interest, should always be included in the model to validate the test on the interaction effect. That is, an interaction effect ~~cannot be interpreted~~ is difficult to interpret in the absence of its main effect (Nelder, 1994; Purcell, 2002).

[Figure 1 about here.]

The example above focused on person-level moderation and whether a latent variable manifests itself differently for different persons (i.e., persons that differ on the person-level moderator). For indicator-level applications, parameter interpretation is similar, however, the key question is whether a latent variable manifests itself differently for different persons-indicator combinations. ~~That is in~~ In indicator-level moderation, a subject may be high on the moderator on one indicator but relatively low on the moderator of the next indicator. For instance, in the intelligence example~~discussed above~~, DiTrapani et al. (2016) found the slope and intercept parameters of a ~~matrix reasoning~~ matrix-reasoning test to be moderated by the item-specific response times. This indicates that the ~~latent matrix reasoning~~ matrix-reasoning latent variable manifests itself differently in the observed indicators for faster responses than for slower responses. Thus, subjects may be high on the moderator of one item (relatively slow response) and low on the moderator on the next item (relatively fast response), see Figure 1 for a graphical representation of the difference between person-level moderation and indicator-level moderation. As the figure shows, the crucial difference between the two modeling frameworks is that in the

13

traditional moderation framework for each person a single value of the moderator is considered, while in indicator-level moderation the moderator varies both across persons and across indicators.

In the case of continuous indicators, the residual variances are ~~an additional parameter~~ additional parameters in the latent variable model. Moderation of the residual variances is generally not considered in tests on interactions, similarly as in regression analysis and path analysis where it is assumed that the residual variances are constant over moderator variables (homoscedasticity). However, in testing for strict measurement invariance, homoscedastic residual variances are a necessary condition. If the residual variances are heteroscedastic, one speaks of weak measurement invariance (Meredith, 1993). Thus, moderation of the residual variances is not strictly necessary in testing for interactions, but it should ideally be considered when establishing strict measurement invariance (see e.g., Bauer, 2017, for a possible approach).

2.2   Moderation by a step function

As discussed ~~above~~in the previous section, an intuitive approach to model the dependence of λ_i and ν_i on Z is to categorize the continuous moderator Z and to fit the latent variable model ~~above~~ to the data of each group ~~that arises~~ using a multi-group approach (e.g., Jöreskog, 1971). That is, the function $f_{\lambda_i}(Z)$ above is given by the following step function

$$f_{\lambda_i}(Z) = \begin{cases} \lambda_{0i} & Z \leq \tau_1 \\ \lambda_{1i} & \tau_1 < Z \leq \tau_2 \\ \ldots & \ldots \\ \lambda_{(C-1)i} & Z > \tau_{C-1} \end{cases} \tag{8}$$

in which τ_c are the thresholds for $c = 1, 2, ..., C-1$ at which the continuous moderator is categorized into C categories. The resulting parameters, λ_{0i}, λ_{1i}, ..., $\lambda_{(C-1)i}$ reflect the slope parameters in the different subgroups. A similar step function can be specified for $f_{\nu_i}(Z)$, resulting in parameters ν_{0i}, ν_{1i}, ..., $\nu_{(C-1)i}$, and for $f_{\ln[\sigma_i]}(Z)$ resulting in parameters $\ln[\sigma_{0i}^2]$, $\ln[\sigma_{1i}^2]$, ..., $\ln[\sigma_{(C-1)i}^2]$.

Similarly as in the multi-group approach to traditional moderation, indicator-level moderation models have been specified by assuming categorized moderators (see Partchev & de Boeck, 2012, and DiTrapani et al., 2016, for a possible approach in the case of dichotomous indicators and dichotomous moderators). The functional relationship between the parameters of the latent variable model and the indicator-specific moderator in this model can be represented similarly as in Equation 8, ~~but~~ by replacing the moderator common for all indicators by the indicator-level moderator, Z_i, and replacing the thresholds by the indicator-specific thresholds, τ_{ci}.

2.3 Moderation by a linear function

A linear moderation model retains the continuous nature of the moderator. Such a model is specified by (see Purcell, 2002; Rabe-Hesketh et al., 2004; Mehta & Neale,

2005; Neale et al., 2006; Bauer & Hussong, 2009; Molenaar et al., 2010):

$$f_{\lambda_i}(Z) = \lambda_{0i} + \lambda_{1i}Z, \tag{9}$$

$$f_{\nu_i}(Z) = \nu_{0i} + \nu_{1i}Z, \tag{10}$$

where ν_{0i} is the intercept parameter of the moderation model and ν_{1i} accounts for the main effect of the moderator in the observed indicators, Y_i. In addition, λ_{0i} is the slope parameter in the moderation model, and λ_{1i} is a moderation parameter which ~~account~~ accounts for the interaction effect between η and Z on Y_i. As discussed by~~for instance~~, for instance, Purcell (2002) and Tucker-Drob (2009), it is straightforward to extend the model above to include higher-order powers of the moderator ~~resulting in~~ with additional parameters λ_{2i}, λ_{3i}, ..., and ν_{2i}, ν_{3i}, ..., depending on the powers of Z that are added (Z^2, Z^3, ...). In addition, in the case of continuous indicators, it is straightforward to introduce a dependency of $\ln(\sigma_i^2)$ on Z as follows:

$$f_{\ln[\sigma_i^2]}(Z) = \beta_{0i} + \beta_{1i}Z. \tag{11}$$

.

Similarly to the traditional linear moderation models, an indicator-level moderation model has been specified by assuming linear functions between the latent variable model parameters and the indicator-level moderators (see Bolsinova, Tijmstra, & Molenaar, 2017; Goldhammer, 2015, for an approach in the case of dichotomous indicators). That is, in Equations 9 and 10 a common moderator Z is replaced with indicator-specific moderator Z_i. Alternatively, a linear model for the log of the slope

16

parameter has also been used (see Bolsinova, de Boeck, & Tijmstra, 2017).

2.4 Moderation by a nonlinear function: parametric and nonparametric approaches

~~In the above, it was mentioned that linear moderation approaches~~ Linear moderation approaches discussed in the previous section do not suffer from the problem of information loss due to categorization, but ~~that~~ they restrict the relationship between the moderator (or its transformation) and the parameters (or their transformations) to be linear. To go beyond these restrictions, the linear moderation model in Equations 9 and 10 ~~have~~ has been extended to include higher-order powers of the moderator resulting in additional parameters λ_{2i}, λ_{3i}, ..., and ν_{2i}, ν_{3i}, ..., depending on the powers of Z that are added (see e.g., Purcell, 2002; Tucker-Drob, 2009; Bauer & Hussong, 2009). However, this has only been done for traditional moderation models, ~~but~~ not for indicator-specific moderators.

Nonlinear parametric models are still rather restrictive since they specify the relationship between the moderator and the parameters of the latent variable model to be a polynomial function. Alternatively, one may want to approach the relationship between the moderator and the parameters from a more flexible and exploratory perspective, which ~~is done in~~ can be done in a local structural equation modeling approach (Hildebrandt et al., 2009; Hülür et al., 2011; Hildebrandt et al., 2016). The idea of local structural equation models is to estimate separate models for different values of the moderator, called focal points. If a moderator is categorical and there are only a few categories, then it boils down to estimating a multi-group model. How-

ever, if a moderator is continuous there will not be enough observations at each focal point to estimate separate models. Instead, observations can be weighted around the focal points and the model parameters can be estimated separately for each focal point on the basis of the weighted sample of the observations. Hence, local structural equation modeling allows exploration of the relationship between the moderator and the parameter of the latent variable model in a very flexible way.

3 A Parametric Nonlinear Approach and a Nonparametric Approach to Indicator-Level Moderation

As appears from the existing approaches ~~discussed above~~previously discussed, the parametric nonlinear and nonparametric moderation approaches are established for moderators that vary across persons but not across indicators. However, in the case of indicator-level moderation, only linear or categorization approaches are available. In this section, we derive a nonlinear parametric approach and a nonparametric approach to indicator-level moderation.

3.1 Parametric nonlinear indicator-level moderation

The assumption of linearity of the moderation effect of the indicator-level moderator can be relaxed by including higher-order effects as follows:

$$f_{\lambda_i}(Z_i) = \lambda_{i0} + \sum_{r=1}^{R} \lambda_{ir} Z_{ir}^*, \tag{12}$$

with the analogous expression for $f_{\nu_i}(Z_i)$ and $f_{\ln[\sigma_i^2]}(Z_i)$, where Z_{i1}^*, $Z_{i2}^*, \ldots, Z_{iR}^*$ are the standardized (i.e., having standard deviation of 1) orthogonal polynomials of degree R over the set of values of the moderator Z_i. We use orthogonal polynomials instead of $\{Z_i, Z_i^2, \ldots, Z_i^R\}$ to facilitate parameter estimation and interpretation. If $R = 1$ the model reduces to an indicator-level linear moderation model, if $R = 2$ the model reduces to a curvilinear indicator-level moderation model, etc. It may be noted that the model in Equation 12 can in principle be seen as a generalized linear model, but we follow Bauer and Hussong (2009) and refer to it as a nonlinear moderation model because it allows for nonlinear relationship between the indicator-specific moderator and the indicator and for nonlinear moderation effects.

3.2 Nonparametric indicator-level moderation

The goal of indicator-level moderation, is for each observed indicator i to explore the relationship between the moderator Z_i and the indicator-specific slope and intercept. Extending the idea of local structural equation modeling, for each observed indicator we define the focal points $F_{i1}, \ldots, F_{iJ}$ of the moderator and obtain the estimates of λ_{ij} and ν_{ij} at each of these focal points by weighting the observations y_{pi} using the

19

distance between the value of the observed moderator z_{pi} of person p and the value of the moderator at the j-th focal point. While in traditional nonparametric moderation each person receives a single weight for each focal point, in indicator-level moderation, weights are assigned to each combination of a person and an indicator separately. Different weighting functions can be used, among which the Gaussian kernel function is rather convenient and intuitive (Gasser, Gervini, Molinari, Hauspie, & Cameron, 2004). The idea behind using the Gaussian kernel is that the observations that are close to each other are more similar than the observations that are further apart. The closer to the focal point, the higher the weight that the observation receives. In this way the estimates of the latent variable model parameters are more influenced by the observations near the focal point and less influenced by the observations further away from it. The relationship between the latent variable model parameters and the indicator-level moderator is approximated by estimating the parameters at each focal point.

For each indicator i and each focal point j a vector of weights $\mathbf{w}_{ij}$ is defined with each element computed as follows:

$$w_{pij} = \exp \left\{ -\frac{(z_{pi} - F_{ij})^2}{2(h\text{SD}_{Z_i} N^{-\frac{1}{5}})^2} \right\}, \tag{13}$$

where w_{pji} is the weight of the observation of person p on indicator i when obtaining the estimates of parameters of i-th indicator at the focal point F_{ij}, z_{pi} is the realisation of the i-th indicator-level moderator for person p, h is the bandwidth factor, and SD_{Z_i} is the standard deviation of the i-th moderator.

The bandwidth factor determines the standard deviation of the normal density

20

function used to weigh the observations, ~~that is, it~~ . It determines how far from the focal points the observation has to be to have a relatively large weight and therefore, a relatively large impact on the estimated model parameters. The bandwidth factor serves as a smoothing parameter: the larger it is, the smoother the relationship between the moderator and the latent variable model parameters is estimated to be. When h is large, the noise in the estimates is reduced, but so is the ability of the method to pick up on the ~~meaninful~~ meaningful but nuanced trends in the data, and the strength of the moderation effect might be underestimated. If $h \rightarrow +\infty$, then the model parameters will be estimated to be the same at each focal point, ~~that is,~~ and there will be no effect of the moderator. The smaller h is, the noisier the estimated relationship between the moderator and the model parameters is. ~~Here~~Hence, the estimates more closely match the local subsets of data, but the chance of picking up on statistical noise is also increased. It must be noted that this trade-off is not specific for nonparametric moderation, but is inherent in kernel regression methodology in general (see e.g., Hart, 2013; Li & Racine, 2007). The factor $h = 1.1$ has been proposed in the nonparametric density estimation literature (see e.g., Silverman, 1986). Alternatively, the value of $h = 2$ has been used in local structural equation modeling (Hildebrandt et al., 2009; Briley et al., 2015). In this paper we use both values and compare the results.

Similar to the traditional nonparametric moderation approach (Hildebrandt et al., 2016) the significance of the moderation effect can be tested using a permutation test (Good, 2005). To perform the permutation test, one needs to estimate the non-parametric relationship between the moderator and the indicator-specific parameters

repeatedly on permuted data sets, ~~that is, on data sets~~ in which the data $\mathbf{y}$ are kept intact, while for each indicator the values of the corresponding moderator are reassigned to different persons in the sample. To make inferences about the significance of the moderation effect on, for example, the slope of the i-th indicator, one needs to compare the standard deviation among the estimates of $\lambda_{i1}, \ldots, \lambda_{iJ}$ obtained in the observed data with the standard deviations of the estimates of the slope parameters in the permuted data sets. The proportion of data sets in which the standard deviation is larger for the permutation data than for the observed data can be used as the p-value for testing the hypothesis of the absence of the moderation effect. Note ~~,~~ that although compared to traditional moderation we have more moderators, we do not estimate more effects than in traditional moderation models and do not perform more permutation tests than Hildebrandt et al. (2016). In addition to the permutation test, one can also use bootstrapping (see e.g., Efron, 1979) to investigate the uncertainty about the estimated relationships between the indicator-level moderator and the model parameters.

4 Estimation

In this section, we describe how the parametric nonlinear and nonparametric relationships between the indicator-specific moderator and the parameters of the latent variable model can be estimated. We take a pragmatic approach when choosing the estimation method: the parametric moderation model is estimated using a Bayesian procedure since it is a model with a large number of parameters which complicates

the use of frequentist methods, and for estimating the nonparametric relationship between the moderator and the model parameters we use maximum likelihood estimation because in this case it is computationally more efficient. We will focus on binary indicator variables as the resulting models are arguably the most challenging models as they are computationally more demanding than the latent variable models for continuous indicators.

4.1   Parametric moderation

To estimate the indicator-level R-degree polynomial moderation 2PNOM we propose using Gibbs sampling (see e.g., Casella & George, 1992; Geman & Geman, 1984). Gibbs sampling allows one to obtain samples from the joint posterior distribution of the model parameters and use these samples to approximate the posterior means of the parameters which can be used as their point estimates and credible intervals which can be used as measures of uncertainty about the parameters.

The joint posterior distribution of the model parameters is proportional to the product of the likelihood and the prior distribution:

$$f(\boldsymbol{\lambda}, \boldsymbol{\nu} \,|\, \mathbf{y}, \mathbf{z}) \propto f(\boldsymbol{\lambda}, \boldsymbol{\nu}) \prod_{p=1}^{N} \int \prod_{i=1}^{K} f(y_{pi} \,|\, \boldsymbol{\lambda}_{i\cdot}, \boldsymbol{\nu}_{i\cdot}, \eta, z_{pi}) \, \mathcal{N}(\eta; 0, 1) d\eta, \qquad (14)$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are the $K \times (R + 1)$ matrices containing the baseline slopes and intercepts, respectively, in the first column and the effects of the moderator on the slopes and the intercepts in the remaining columns, respectively. Here, y_{pi} ~~s are~~ is binary and the conditional probability of $y_{pi} = 1$ given η is specified by the 2PNOM.

23

As a prior distribution for the indicator-specific sets of parameters we use a hierarchical multivariate normal prior:

$$f(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \int \int \prod_{i=1}^{K} \mathcal{N}_{2R+2}([\boldsymbol{\lambda}_{i\cdot}, \boldsymbol{\nu}_{i\cdot}]^T; \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(\boldsymbol{\mu}) f(\boldsymbol{\Sigma}) d\boldsymbol{\mu} d\boldsymbol{\Sigma}, \tag{15}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the hyper-mean and hyper-covariance matrix. Using a hierarchical prior allows one to investigate the relationship between the different model parameters since the hyper-parameters can be also estimated. For example, using the posterior distribution of the covariance between $\boldsymbol{\lambda}_{\cdot 0}$ and $\boldsymbol{\lambda}_{\cdot 1}$ one can make inferences about the relationship between the baseline slope and the linear moderation effect. For the the hyper-mean and the hyper-covariance we use a vague multivariate normal and a vague inverse-Wishart prior, respectively.

In the Gibbs sampler after setting initial values for the model parameters, they are consecutively sampled from their full conditional posterior distributions given the current values of all other parameters. The conditional posteriors for the distribution in Equation 14 do not have a closed form. To simplify the conditional posteriors data augmentation (Tanner & Wong, 1987) is implemented. First, for each binary data point y_{pi} an augmented continuous variable $x_{pi} \sim \mathcal{N}((\boldsymbol{\lambda}_{i\cdot}^T \mathbf{z}_p^*) \eta_p + \boldsymbol{\nu}_{i\cdot}^T \mathbf{z}_p^*), 1)$ is introduced (Albert, 1992), defined in such a way that $y_{pi} = \mathcal{I}(x_{pi} \geq 0)$. Second, parameters η_p of each person are sampled. Third, the hyper-parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are also sampled. Hence, in the data augmented Gibbs Sampler samples are obtained from the following joint posterior distribution:

$$p(\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{x}, \,|\, \mathbf{y}, \mathbf{z}). \tag{16}$$

All full conditional posteriors of this joint posterior have a closed form: truncated normal for x_{pi}s, normal for η_ps, multivariate normals for $[\boldsymbol{\lambda}_{i\cdot}, \boldsymbol{\nu}_{i\cdot}]^T$ and $\boldsymbol{\mu}$, and Inverse-Wishart for $\boldsymbol{\Sigma}$. The exact specification of initial values and the steps of the Gibbs Sampler can be found in the Appendix. R and C code for estimating the R-degree polynomial indicator-level moderation model can be found in the supplementary materials.

4.2   Nonparametric moderation

While in traditional nonparametric moderation the model is estimated once for each focal point, when the moderator is indicator-specific the procedure needs to be iteratively repeated for each indicator. To start the procedure one needs to initialize the values of the parameters for each combination of a person and an indicator, that is, define the $N \times K$ matrices of person-by-indicator-specific slopes and intercepts, denoted by $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$, respectively. The initial values can be set as follows: estimate the model without the moderation effects and set each λ_{pi}^* and ν_{pi}^* to be equal to the estimates of the slope and the intercept of the i-th indicator in the model without the moderation effects.

After initialisation, one first repeatedly obtains the estimates of λ_{ji} and ν_{ji} for $i \in [1:K]$ and for $j \in [1:J]$ by maximizing the log-likelihood:

$$\ln \mathcal{L}(\lambda_{ij}, \nu_{ij}\,;\, \mathbf{y}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*, \mathbf{w}_{ij}) =$$

$$\sum_{p=1}^{N} \ln \int f(y_{pi}; \lambda_{ij}, \nu_{ij}, \eta)^{w_{pij}} \prod_{k \neq i, k=1}^{K} f(y_{pk}; \lambda_{pk}^*, \nu_{pk}^*, \eta) \mathcal{N}(\eta; 0, 1) d\eta \quad (17)$$

where the observations of indicator i are weighed by $\mathbf{w}_{ij}$, while for the indicators $k \neq i$ the current values of the person-by-indicator-specific slope and intercept parameters for the k-th indicator, $\boldsymbol{\lambda}^*_{\cdot k}$ and $\boldsymbol{\nu}^*_{\cdot k}$, are used. In this study we used numerical integration to approximate the integral in Equation 17 (Gauss-Hermite quadrature with six nodes) and general-purpose optimisation R-function 'optim' to find the values of λ_{ij} and ν_{ij} that maximize the log-likelihood.

Next, one updates the values of $\boldsymbol{\lambda}^*_{\cdot i}$ and $\boldsymbol{\nu}^*_{\cdot i}$ as follows:

$$\lambda^*_{pi} = \begin{cases} \lambda_{i1} \text{ if } z_{pi} < F_{i1}, \\[2mm] \lambda_{ij} + (z_{pi} - F_{ij})\frac{\lambda_{i(j+1)} - \lambda_{ij}}{F_{i(j+1)} - F_{ij}} \text{ if } F_{ij} \leq z_{pi} \leq F_{i(j+1)}, \forall j \in [1, J-1], \\[2mm] \lambda_{iJ} \text{ if } z_{pi} > F_{iJ}; \end{cases} \tag{18}$$

with a similar specification for ν^*_{pi}. ~~That is, if~~ If z_{pi} is outside of the range of the focal points, then λ^*_{pi} and ν^*_{pi} are set to be equal to the parameters at the nearest focal point, and if z_{pi} is between F_{ij} and $F_{i(j+1)}$, then λ^*_{pi} and ν^*_{pi} are computed using piece-wise linear regression. The values outside the range of the focal points are set to be equal to the parameters at F_{i1} and F_{iJ} because if one uses the regression-based values, extremely low ($\rightarrow -\infty$) or extremely high ($\rightarrow +\infty$) values may be obtained which may be unrealistic in practical applications. Given the results of the simulation studies, piece-wise linear approximation was considered accurate enough~~, however~~. However alternatively one may consider fitting a higher degree polynomial regression and use the values on the smooth curve connecting the estimates at the focal points for λ^*_{pi} and ν^*_{pi}.

The procedure needs to be repeated more than once to get rid of the effect of the starting values. However, our simulations have shown that the estimates of $\lambda_{i1}, \ldots, \lambda_{iJ}$ and $\nu_{i1}, \ldots, \nu_{iJ}$ are already stable after the second repetition. R and C code for the indicator-level nonparametric moderation can be found in the supplementary materials.

5   Illustrative exampleExample

5.1   Data and method

To illustrate the differences between the moderation methods we used the data from a computerized arithmetic test which was part of the central exams in the Netherlands at the end of secondary education. The same data were used by Bolsinova and Maris (2016). From the data of a test version with 60 items administered to 10,367 persons we removed the last 10 items which many respondents did not reach (this was done since the response behavior under strong time pressure might be different from the rest of the test) and removed all persons with missing values on the remaining 50 items.

The resulting data set consisted of responses from 9,697 persons to 50 arithmetic items. Here, y_{pi} is the binary response accuracy (1 - correct, 0 - incorrect) of person p to item i. In addition to the response accuracy, the data set included item-level response times which were used as a moderator for illustrating the moderation methods (i.e., here, z_{pi} is response time of person p on item i). The 2PNOM with the standard normal distribution for the latent variable (i.e., arithmetic ability) was used

in this application.

The 2PNOM without moderation effects, the linear moderation model, the parametric nonlinear models (quadratic and cubic) and the model for the dichotomized moderator (split at the item-level median) were fit to the data using Gibbs Samplers with 10,000 iterations each (including 5,000 of burnin). Modified AIC and BIC[2] (mAIC abnd mBIC) were used for model comparison.

Furthermore, the nonparametric moderation method for investigating the relationship between response time and the slope and intercept parameters ~~were~~ was applied to the data. For each item in the nonparametric method, we used 20 equally spaced percentiles (from the 2.5th to the 97.5th percentile) of the distribution of the response times as focal points. Additionally, we performed the permutation test with 100 replications to evaluate the significance of the moderation effects. Moreover, we used 100 bootstrap samples (each time 9,697 persons were sampled with replacement from the available sample) to graphically evaluate the uncertainty about the estimated relationships. The above described procedure was performed twice: with $h = 1.1$ and $h = 2$.

5.2   Results

[Figure 2 about here.]

Convergence of the models was evaluated using visual inspection of the trace

---

[2]The information criteria are referred to as *modified* since instead of the maximum likelihood estimates of the parameters, the posterior means of the parameters were used. The utility of these criteria for model selection of linear indicator-level moderation models was demonstrated by Bolsinova et al. (2017).

plots and the running mean plots, which did not provide evidence for the absence of convergence. As an example, Figure 2 shows for the most complex model (cubic moderation) the plots for the parameters of item 10. The black lines show the values of each parameter throughout the iterations and the red lines show the running means. It can be seen that the running means stabilize after the burn-in and the sampled parameter values fluctuate around these means covering the range of the posterior distribution.

~~In Table~~ To further assess convergence we ran five additional chains with more iterations (20,000 each including 5,000 burn-in) for each model with varying starting values and assessed convergence using the Gelman-Rubin diagnostic (Gelman & Rubin, 1992). For the model without a moderation effect, for the linear moderation model, and for the model with the dichotomized moderator, all parameters showed good convergence. In the quadratic and the cubic models the convergence diagnostics indicated that the item parameters of 47 and 48 items, respectively, showed good convergence. For the remaining three and two items the diagnostics indicated some convergence issues, which upon inspection showed that for both models two different sets of solutions were found across chains with three out of five chains stuck in a local maximum. When convergence was evaluated for the two chains which were not stuck in the local maximum all item parameters did not show lack of convergence. For both models, the results found for their original chain matched the results of the respective best chains. Thus, there do not appear to be any convergence issues in the original results, and hence for each model we proceeded with the analysis of the results of the original chain.

[Table 1 about here.]

[Figure 3 about here.]

In Table 1 the models with and without the moderation effects are compared in terms of the information criteria. All models with moderation effects performed better than the model without the moderation effects which indicates that in this application, response accuracy is not independent of response time. The linear model performed better than the model with a dichotomized moderator. However, the assumption of linearity of all moderation effects can be rejected since the models with higher-order effects have smaller mAIC and mBIC than the linear model. The model-based uncertainty about the estimated functional relationship between the moderator and the model parameters can be evaluated by inspecting the posterior distributions of the moderation effects. Figure 3 shows examples for some of the parameters (slopes of items 5 and 7, and intercepts of items 3 and 27). Each grey line is based on one sample from the posterior distribution of model parameters, and the black line is based on the posterior means.

[Figure 4 about here.]

[Figure 5 about here.]

The conclusion of the presence of the moderation effects also follows from the permutation tests performed with the nonparametric moderation method. Among the p-values of the permutation tests, 48 and 49 of them were significant for the effect on the intercept parameter for $h = 1.1$ and $h = 2$, respectively, indicating a significant

30

main effect of Z_i on Y_i. In addition, 33 and 35 tests were significant for the effect on the slope parameter for $h = 1.1$ and $h = 2$, respectively, indicating a significant interaction effect between Z_i and η. As an illustration ~~ofddd~~ of the permutation test, Figure 4 shows the relationship between response time and the parameters of the latent variable model for some of the items. The effect was significant, for example, for the slope parameter of item 5 and the intercept parameter of item 3. On the contrary, the effect on the slope parameter of item 7 and the effect on the intercept parameter of item 27 were not significant. For these items, Figure 5 also shows the estimated relationships in the bootstrap samples. The estimated relationship for λ_7 and ν_{27} in the bootstrap samples fluctuates around a horizontal line. One can also see that the uncertainty about the estimated relationship is larger when $h = 1.1$, compared to $h = 2$, and that the estimated effect is stronger with $h = 1.1$.

[Figure 6 about here.]

[Figure 7 about here.]

Figures 6 and 7 show examples of the estimated relationship between the response time and the slope and the intercept parameters as obtained from the different methods. For some items there are no large differences between the linear model and the nonlinear models (see the slope parameter of item 45 and the intercept parameter of item 5). However, for other items there is a large discrepancy between the linear and nonlinear models (see the slope parameter of item 34 and the intercept parameter of item 48). The slope of item 34 first increases with response time and then decreases. The quadratic and the cubic ~~model~~ models accommodate this change of the direction of the effect, but they do not closely match the relationship estimated with the

31

nonparametric method. The intercept parameter of item 48 first increases steeply but there seems to be hardly any difference in the intercept parameters given the response times above 200 seconds. For this effect the parametric nonlinear models give a rather good approximation. Furthermore, one can see from the figures that the assumption of the stability of the parameters within the categories of the moderator does not seem to hold. If we compare the results for the two different values of h, we see that the general patterns of the estimated relationships are very close to each other, but with $h = 2$ the effect is weakened and small differences between the neighboring focal points are ~~ignored~~smoothed out.

6   Simulation ~~study~~Study

6.1   Methods

In this simulation study the viability of the parametric nonlinear and nonparametric methods is investigated. In order to obtain simulated data under a realistic relationship between the latent variable model parameters and the continuous moderator, we used the moderator values and estimates of the person-by-indicator-specific slopes and intercepts from the illustrative example. In the simulation study we again used the latent variable model for binary indicators, specifically the 2PNOM to illustrate the viability of nonlinear moderation.

From the empirical data set with 50 indicators we selected 24 indicators with different patterns for the relationship between the indicator-specific moderator and the slope and intercept parameters. Additionally, we considered an indicator with

32

no moderator effect on the parameters of the latent variable model, ~~that is,~~ with a constant slope parameter of .58 (i.e., matching a slope of 1 in the two-parameter logistic model) and a constant intercept equal to 0 for which we used the values of the moderator of one of the remaining 26 indicators in the arithmetic test.

Two scenarios were considered: 1) the data were generated using the cubic moderation model; 2) the data follow a latent variable model with person-by-indicator-specific slope and intercept parameters, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^{*3}$. In both scenarios, the true parameter values were equal to the parameter estimates of the selected 24 indicator variables in the real data illustration ~~(as discussed above)~~ in the cubic moderation model and the nonparametric moderation method respectively.

In each replication, N persons were drawn from the available data set of 9657 persons and their response times were used in the simulation as the values of the indicator-specific moderator. In the first scenario, for each combination of a person and an indicator, we computed the slope and the intercept parameters using the corresponding value of the moderator and the estimates of the parameters of the cubic moderation model. In the second scenario, the estimates of the person-by-indicator-specific slope and intercept parameters, λ_{pi}^* and ν_{pi}^*, of the selected N persons obtained with the nonparametric method were used to generate the data of the indicator variables. For each person, a value for the latent variable was drawn from the standard normal distribution. The values of the indicators were generated according to the 2PNOM.

Three conditions with different N were considered: 1000, 2000, and 4000 persons.

---

[3]We used the results with the bandwidth $h = 1.1$

In each ~~generated~~ condition 100 data sets were generated. In each generated data set we estimated a cubic moderation model (a Gibbs Sampler with 10,000 iterations and 5,000 burn-in was used, see Appendix for details) and the relationship between the parameters of the latent variable model and the indicator-specific moderator was investigated using the nonparametric method, both with $h = 1.1$ and $h = 2$. As focal points we used the 20 equally spaced percentiles (from the 2.5th to the 97.5th) of the distribution of the indicator-specific moderator in the full data set.

The goal of the simulation study was to investigate the recovery of the true relationship between the indicator-specific moderator and the parameters of the latent variable model. The estimates of the focal point-specific slope and intercept parameters were compared to their true values. For each indicator we computed the weighted absolute bias, variance and mean squared error of the indicator-specific parameters at the 20 focal points, separately for the slope and for the intercept parameters. The weights for each focal point were determined by the kernel density estimates at that point.

6.2   Results

[Table 2 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

34

Table 2 shows the results for recovery of the indicator-specific parameters averaged across the indicators. Additionally, Figures 8, 9, 10, and 11 show examples of how the estimates differ across the estimation methods and the sample sizes. One can see that generally the estimated curves follow the patterns of the true relationship. Furthermore, increasing sample size helped reduce variability of the estimated curves.

When the data were generated under the cubic model, the parametric method performed better than the nonparametric method. The bias was small and can probably be at least partly explained by the shrinkage effect, since the hierarchical prior was used for the indicator-specific parameters. The variance was also smaller than that under the nonparametric method, which is not surprising, since the cubic model specifies the parametric shape of the relationship between the parameters of the latent variable model and the moderator and therefore restricts the range of possible functions. With the nonparametric method the variance was larger, since it allows for more flexibility in the shape of the moderation effect. Furthermore, bias was larger with the nonparametric method. When estimating λ_{ij} and ν_{ij} at a particular focal point, observations removed from the focal point also influence the results, therefore there was less variation in the estimated focal point-specific parameters than in the true ones, this. This effect is more pronounced for $h = 2$ than for $h = 1.1$. With $h = 2$ strong effects are heavily underestimated. For example, in Figure 8 in the bottom row , which represents of the results of the nonparametric method, one can see that at the lowest and the highest focal points, the intercepts were overestimated, while in the middle they were underestimated. Overall, the

35

~~effect of the moderator is therefore slightly underestimated since the decrease of the intercept when moving away from the median of the moderator is estimated to be less strong~~ Thus, the strength of the moderation effect is underestimated by this method. The difference between the results with $h = 1.1$ and $h = 2$ depends on the curvature of the moderation function. For example, in Table 2 one can see that the difference between the bias with $h = 1.1$ and $h = 2$ was larger for the intercepts, than for the slopes, since the moderation effect were generally stronger for the intercepts.

When the data were generated using the person-by-indicator-specific slopes and intercepts estimated with the nonparametric method, the benefits of the cubic model become less prominent. The variance of the estimates was smaller under the cubic model, however, for those indicators for which a cubic regression was not a good approximation of the underlying relationship between the moderator and the parameters of the latent variable model, the bias under the parametric model was larger than the bias under the nonparametric method. Figures 10 and 11 show that while the parametric model forces a particular shape on the relationship between the moderator and the parameters of the latent variable model and, hence, cannot recover the true shape of the relationship, the relationship estimated with the nonparametric method very closely matches the true relationship for $h = 1.1$. With larger sample size the difference in the variance of the estimates under the parametric and the nonparametric methods becomes smaller~~and the~~. Since the nonparametric method has smaller bias, the overall quality of the estimates ~~, in terms of~~ as captures by the mean squared error ~~, points in the direction of better performance of the nonparametric method~~ is better for this method compared to the parametric method when sample

size is large. In conclusion, when the true relationship deviates from the cubic model and the sample size is large, it is more beneficial to use a more flexible nonparametric approach, which has larger variance, but also smaller bias.

7   Discussion

The methods presented in this paper allow researchers to move beyond the assumptions of linearity of the moderation effects and homogeneity of the indicator-specific parameters of the latent variable model within artificially created categories of the moderator. As our illustrative example shows, it is not uncommon for the relationship between the indicator-specific moderator and the parameters to be nonlinear. Furthermore, while sometimes a quadratic or a cubic model can give a good approximation of the moderation effect, in other cases the patterns of the relationship may not be accurately represented by these parametric models. Therefore, having an exploratory nonparametric tool for studying the relationship between the indicator-specific moderator and the parameters of the latent variable model is very useful. The estimated nonparametric relationship might be less strong than the true relationship, but the bias of the nonparametric method would be smaller than the bias in the parametric model when the shape of the true relationship is far from the specified parametric shape.

The nonparametric approach allows one to investigate the relationship between the moderators and the parameters of the latent variable model in an exploratory way and use what has been learned in this explanatory step in further research. The

37

results of the nonparametric method can be used to inform the parametric model in further confirmatory studies about what degree polynomial is needed to approximate the relationship between the indicator-specific moderator and the parameters of the model.

When specifying the weights for the nonparametric estimation of the relationship between the indicator-specific moderator and the parameters, in this study a normal density kernel was used, which assumes that observations, for example, one unit below and one unit above the focal point are equally similar to the observations at the focal point. Alternatively, one might assume that observations with values, for example, twice as large or twice as small as the value of the focal point are equally similar to the observations at the focal point. In that case, it would be better to use a log-transformation for the indicator-specific moderator. Choosing the method for specifying the weights depends on the application at hand.

When choosing the value for the bandwidth factor we followed recommendations from the nonparametric regression literature and from local structural equation modeling. Both $h = 1.1$ and $h = 2$ were used in the empirical data analysis. The results show that although the results are smoother with $h = 2$, the general trends of the relationship are similar regardless of the exact value of h. Furthermore, the results of the permutation tests are also very similar. As a general recommendation for users of the method we suggest different values of h and checking whether the same conclusions would be made about general patterns of relationship between the indicator-level moderator and the parameters of the latent variable models are the same. While the exact values of ν_{ij} and λ_{ij} are sensitive to the exact choice of h (see

for the example the results of the simulation study, where the MSE of nonparametric method was different with different h), the general pattern of relationship which is of primary interest is often similar for $h = 1.1$ and $h = 2$. An alternative to using fixed values of h would be adapting one of the data-driven methods for bandwidth selection from nonparametric regression and kernel density estimation literature (see e.g., Bowman, 1984; Chiu, 1991, 1992; Hall, Sheather, Jones, & Marron, 1991; Hall, Marron, & Park, 1992; Park & Marron, 1990; Rice, 1984; Rudemo, 1982; Scott & Terrell, 1987; Sheather & Jones, 1991). However, in the case of indicator-level moderation it is computationally very expensive to compute the exact predicted value of the latent variable model parameters for each combination of the person with the indicator which complicates the application of these methods. Further research is needed to develop data-driven methods for optimal selection of the bandwidth factor h that would be suitable for nonlinear moderated latent variable models, not only for the indicator-level moderation discussed in this paper, but also for traditional nonparameteric moderation (Hildebrandt et al., 2009; Briley et al., 2015).

In this paper we used the latent variable model for binary indicators in the illustrative example and the simulation study. However, our methods can be also be adapted to be used in the context of continuous indicators or ordinal indicators. Furthermore, while in this study we considered a unidimensional latent variable model in which the conditional expectation of the observed indicator is modeled with an intercept and a single slope parameter, the methods described here can be readily extended to more complex multidimensional latent variable models and structural equation models.

Appendix: Gibbs Sampler for the indicator-level R-degree polynomial moderation 2PNOM.

As starting values we use maximum likelihood estimates of the slopes and intercepts in the non-moderated model for $\boldsymbol{\lambda}_{\cdot 0}$ and $\boldsymbol{\nu}_{\cdot 0}$, respectively, EAP-estimates of the person parameters for η_ps, $\mathbf{0}$ for all the effect parameters (e.g., $\boldsymbol{\lambda}_{\cdot 1}$) and for $\boldsymbol{\mu}$, and an identity matrix for $\boldsymbol{\Sigma}$. Initial values do not have to be specified for x_{pi}s since they are sampled in the first step.

After initialisation, the parameters (and the augmented data) are consecutively sampled from their full conditional posterior distributions which are specified in the steps below:

Step 1. For each combination of person p with indicator i sample the augmented variable x_{pi} from its full conditional posterior which is a truncated normal distribution where the truncation depends on the value of y_{pi} (below zero for correct and above zero for incorrect):

$$f(x_{pi} \,|\, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{z}) = f(x_{pi} \,|\, \boldsymbol{\lambda}_{i\cdot}, \boldsymbol{\nu}_{i\cdot}, \eta_p, y_{pi}, z_{pi}) =$$

$$\mathcal{N}(x_{pi}; (\boldsymbol{\lambda}_{i\cdot}^T \mathbf{z}_{pi}^*)\eta_p + \boldsymbol{\nu}_{i\cdot}^T \mathbf{z}_{pi}^*), 1)(y_{pi}\mathcal{I}(x_{pi} \geq 0) + (1 - y_{pi})\mathcal{I}(x_{pi} < 0)). \quad (19)$$

Note that given the matrix of augmented data $\mathbf{x}$ the model parameters are independent of the data $\mathbf{y}$.

Step 2. For each person p sample η_p from its full conditional posterior which is a

normal distribution:

$$f(\eta_p \,|\, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{x}, \mathbf{z}) = f(\eta_p \,|\, \boldsymbol{\lambda}, \boldsymbol{\nu}, \mathbf{y}_{p\cdot}, \mathbf{z}_{p\cdot})$$

$$= \mathcal{N}\left(\eta_p; \frac{\sum_{i=1}^{K} \boldsymbol{\lambda}_{i\cdot}^{T} \mathbf{z}_{pi}^{*}(x_{pi} - \boldsymbol{\nu}_{i\cdot}^{T} \mathbf{z}_{pi}^{*})}{\sum_{i=1}^{K} \boldsymbol{\lambda}_{i}^{T} \mathbf{z}_{pi}^{*} + 1}, 1\right) \quad (20)$$

Step 3. For each indicator i sample the parameters $[\boldsymbol{\lambda}_{i\cdot}, \boldsymbol{\nu}_{i\cdot},]^{T}$ from a multivariate normal distribution with posterior covariance matrix

$$\boldsymbol{\Omega}_i = \left(\mathbf{v}^{T}\mathbf{v} + \boldsymbol{\Sigma}^{-1}\right)^{-1}, \quad (21)$$

where $\mathbf{v} = \begin{bmatrix} \boldsymbol{\eta} & \boldsymbol{\eta} \circ \mathbf{z}_{\cdot i1}^{*} & \boldsymbol{\eta} \circ \mathbf{z}_{\cdot i2}^{*} & \boldsymbol{\eta} \circ \mathbf{z}_{\cdot i3}^{*} & \mathbf{1} & \mathbf{z}_{\cdot i1}^{*} & \mathbf{z}_{\cdot i2}^{*} & \mathbf{z}_{\cdot i3}^{*} \end{bmatrix}$, and posterior mean vector

$$\boldsymbol{\zeta}_i = \boldsymbol{\Omega}_i \left(\mathbf{v}^{T}\mathbf{x}_{\cdot i} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right). \quad (22)$$

Step 4. Sample the covariance matrix of the vector of parameters:

$$\boldsymbol{\Sigma} \sim \mathcal{IW}\left(K + 2R + 4, \mathbf{I}_{2R+2} + \sum_{i=1}^{K}\left(\begin{bmatrix} \boldsymbol{\lambda}_{i\cdot} & \boldsymbol{\nu}_{i\cdot} \end{bmatrix}^{T} - \boldsymbol{\mu}\right)\left(\begin{bmatrix} \boldsymbol{\lambda}_{i\cdot} & \boldsymbol{\nu}_{i\cdot} \end{bmatrix} - \boldsymbol{\mu}^{T}\right)\right).$$
$$(23)$$

Step 5. Sample the mean vector of the parameters:

$$\boldsymbol{\mu} \sim \mathcal{N}_8\left(\left(K\boldsymbol{\Sigma}^{-1} + \frac{\mathbf{I}_8}{100}\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{1}^{T}\begin{bmatrix} \boldsymbol{\lambda}_{i\cdot} & \boldsymbol{\nu}_{i\cdot} \end{bmatrix}\right), \left(K\boldsymbol{\Sigma}^{-1} + \frac{\mathbf{I}_8}{100}\right)^{-1}\right). \quad (24)$$

References

Albert, J. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, *17*, 251-269. doi: 10.2307/1165149

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological methods*, *22*(3), 507. doi: 10.1037/met0000077

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological methods*, *14*(2), 101. doi: 10.1037/a0015583

Bolsinova, M., de Boeck, P., & Tijmstra, J. (2017). Modeling conditional dependence between response time and accuracy. *Psychometrika*, *82*(4), 1126–1148. doi: 10.1007/s11336-016-9537-6

Bolsinova, M., & Maris, G. (2016). A test for conditional independence between response time and accuracy. *British Journal of Mathematical and Statistical Psychology*, *69*, 62-79. doi: 10.1111/bmsp.12059

Bolsinova, M., Tijmstra, J., & Molenaar, D. (2017). Response moderation models for conditional dependence between response time and response accuracy. *British Journal of Mathematical and Statistical Psychology*. doi: 10.1111/bmsp.12076

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, *71*(2), 353–360. doi: 10.2307/2336252

Briley, D. A., Harden, K. P., Bates, T. C., & Tucker-Drob, E. M. (2015). Nonparametric estimates of gene$\times$ environment interaction using local struc-

tural equation modeling. *Behavior genetics*, *45*(5), 581–596. doi: 10.1007/s10519-015-9732-8

Casella, G., & George, E. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*, 167-174. doi: 10.2307/2685208

Chiu, S.-T. (1991). Bandwidth selection for kernel density estimation. *The Annals of Statistics*, 1883–1905. doi: 10.2307/2241909

Chiu, S.-T. (1992). An automatic bandwidth selector for kernel density estimation. *Biometrika*, *79*(4), 771–782. doi: 10.1093/biomet/79.4.771

Cohen, J. (1983). The cost of dichotomization. *Applied psychological measurement*, *7*(3), 249–253. doi: 10.1177/014662168300700301

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., . . . Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate behavioral research*, *49*(3), 214–231. doi: 10.1080/00273171.2014.889594

DiTrapani, J., Jeon, M., De Boeck, P., & Partchev, I. (2016). Attempting to differentiate fast and slow intelligence: Using generalized item response trees to examine the role of speed on intelligence tests. *Intelligence*, *56*, 82 - 92. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0160289616300848` doi: http://dx.doi.org/10.1016/j.intell.2016.02.012

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*, 1-26. doi: 10.1214/aos/1176344552

Facon, B. (2006). Does age moderate the effect of IQ on the differentiation of cognitive abilities during childhood? *Intelligence*, *34*(4), 375–386. doi: 10

.1016/j.intell.2005.12.003

Fox, J. (2015). *Applied regression analysis and generalized linear models.* Sage Publications.

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist*, *1*(9), 372. doi: 10.1037/h0056380

Gasser, T., Gervini, D., Molinari, L., Hauspie, R., & Cameron, N. (2004). Methods in human growth research. *Cambridge Studies in Biological and Evolutionary Anthropology*(39), 179–204.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457-472. doi: 10.1214/ss/1177011136

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741. doi: 10.1109/TPAMI.1984.4767596

Goldhammer, F. (2015). Measuring ability, speed, or both? challenges, psychometric solutions, and what can be gained from experimantal control. *Measurement*, *13*, 133-164. doi: 10.1080/15366367.2015.1100020

Goldhammer, F., Steinwascher, M. A., Kroehne, U., & Naumann, J. (2017). Modelling individual response time effects between and within experimental speed conditions: A glmm approach for speeded tests. *British Journal of Mathematical and Statistical Psychology*, *70*(2), 238–256. Retrieved from `http://dx.doi.org/10.1111/bmsp.12099` doi: 10.1111/bmsp.12099

Good, P. I. (2005). *Permutation, parametric and bootstrap tests of hypotheses: a practical guide to resampling methods for testing hypotheses.* New York:

Springer-Verlag.

Gvozdenko, E. (2010). *Meaning and potential of test response time and certainty data: Teaching perspective.* (Unpublished doctoral dissertation). The University of Melbourne.

Hall, P., Marron, J., & Park, B. U. (1992). Smoothed cross-validation. *Probability theory and related fields*, *92*(1), 1–20. doi: 10.1007/BF01205233

Hall, P., Sheather, S. J., Jones, M., & Marron, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, *78*(2), 263–269. doi: 10.2307/2337251

Hart, J. (2013). *Nonparametric smoothing and lack-of-fit tests.* Springer Science & Business Media. doi: 10.1007/978-1-4757-2722-7

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research*, *51*(2-3), 257-258. Retrieved from `http://dx.doi.org/10.1080/00273171.2016.1142856` (PMID: 27049892) doi: 10.1080/00273171.2016.1142856

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, *16*(2), 87–102.

Hülür, G., Wilhelm, O., & Robitzsch, A. (2011). Intelligence differentiation in early childhood. *Journal of Individual Differences*. doi: 10.1027/1614-0001/a000049

Jeon, M., De Boeck, P., & van der Linden, W. (2017). Modeling answer change behavior: An application of a generalized item response tree model. *Jour-*

nal of Educational and Behavioral Statistics, 42(4), 467–490. doi: 10.3102/
1076998616688015

Jöreskog, K. G. (1971, Dec 01). Simultaneous factor analysis in several populations.
Psychometrika, 36(4), 409–426. Retrieved from `https://doi.org/10.1007/`
`BF02291366` doi: 10.1007/BF02291366

Lee, S.-Y., Poon, W.-Y., & Bentler, P. (1989). Simultaneous analysis of multivariate
polytomous variates in several groups. Psychometrika, 54(1), 63–73. doi:
10.1007/BF02294449

Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: theory and practice.
Princeton University Press.

Lord, F. (1980). Applications of item response theory to practical testing problems.
Hillsdale, N.J.: Erlbaum.

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

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the
practice of dichotomization of quantitative variables. Psychological methods,
7(1), 19. doi: 10.1037//1082-989X.7.1.19

McClelland, G., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015).
Median splits, type II errors, and false positive consumer psychology: Don't
fight the power. Journal of Consumer Psychology, 25. doi: 10.1016/j.jcps.2015
.05.006

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural
equations modeling. Psychological methods, 10(3), 259-284. doi: 10.1037/

1082-989X.10.3.259

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*, 127-143. doi: 10.1016/0883-0355(89)90002-5

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin*, *115*(2), 300. doi: 10.1037/0033-2909.115.2.300

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, *29*(2), 177–185. doi: 10.1007/BF02289699

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi: 10.1007/BF02294825

Millsap, R. (2011). *Statistical approaches to measurement invariance.* New-York: Routledge. doi: 10.4324/9780203821961

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*(6), 611–624. doi: 10.1016/j.intell .2010.09.002

Murray, A. L., Booth, T., & Molenaar, D. (2016). Personality differentiation by cognitive ability: An application of the moderated factor model. *Personality and Individual Differences*, *100*, 73–78. doi: 10.1016/j.paid.2016.03.094

Muthen, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, *46*(4), 407–419. doi: 10.1007/BF02293798

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*(4), 557–585. doi: 10.1007/BF02296397

Neale, M. C., Aggen, S. H., Maes, H. H., Kubarych, T. S., & Schmitt, J. E. (2006). Methodological issues in the assessment of substance use phenotypes. *Addictive Behaviors*, *31*(6), 1010–1034. doi: 10.1016/j.addbeh.2006.03.047

Nelder, J. A. (1994). The statistics of linear models: back to basics. *Statistics and Computing*, *4*(4), 221–234. doi: 10.1007/BF00156745

Park, B. U., & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, *85*(409), 66–72. doi: 10.2307/2289526

Partchev, I., & de Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence*, *40*, 23-32. doi: 10.1016/j.intell.2011.11.002

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research and Human Genetics*, *5*(6), 554–571. doi: 10.1375/136905202762342026

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167–190. doi: 10.1007/BF02295939

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 1215–1230. doi: 10.1214/aos/1176346788

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, *25*(1), 127–141. doi: 10.1002/sim.2331

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 65–78.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph, No. 17*. doi: 10.1002/j.2333-8504.1968 .tb00153.x

Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the american Statistical association*, *82*(400), 1131–1146. doi: 10.2307/2289391

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 683–690. doi: 10.2307/2345597

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of of Psychology*, *15*, 72-101. doi: 10.2307/1412159

Tanner, M., & Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*, 528-540. doi: 10.2307/2289457

Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, *99*, 118-128. doi: 10.1037/ 0033-2909.99.1.118

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental psychology*, *45*(4), 1097. doi: 10.1037/a0015864

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psycholog-*

ical science, *14* (6), 623–628. doi: 10.1046/j.0956-7976.2003.psci_1475.x

Wu, H., & Zhang, J.-T. (2006). *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches* (Vol. 515). John Wiley & Sons. doi: 10.1002/0470009675

Moderated latent variable models Traditional moderation Indicator-level moderationMulti-group

Jöreskog (1971) Partchev and de Boeck (2012)Muthen and Christoffersson (1981) DiTrapani et al. (20

Rabe-Hesketh et al. (2004) Bolsinova, Tijmstra, and Molenaar (2017)Mehta and Neale (2005)Goldhar

parametric Purcell (2002) -Tucker-Drob (2009) Nonparametric Hildebrandt et al. (2009) -Hülür et al.

Table 1: Information criteria for the fitted models: modified AIC (mAIC) and modified BIC (mBIC)

Model	# parameters	mAIC	mBIC
No moderation	100	517569.2	518287.2
Dichotomised moderator	200	509162.2	510598.1
Linear moderation	200	510277.4	511713.3
Quadratic moderation	300	504072.8	506226.7
Cubic moderation	400	500613.8	503485.6

Table 2: Parameter recovery

| Scenario	Method	N	Slope			Intercept		
			Bias	Var	MSE	Bias	Var	MSE
1	Parametric	1000	0.015	0.008	0.008	0.010	0.006	0.006
		2000	0.008	0.004	0.004	0.006	0.003	0.003
		4000	0.005	0.002	0.002	0.003	0.002	0.002
	Non-parameteric	1000	0.033	0.012	0.014	0.033	0.008	0.011
	$h = 1.1$	2000	0.026	0.007	0.008	0.029	0.004	0.007
		4000	0.025	0.004	0.005	0.027	0.002	0.004
	Non-parameteric	1000	0.041	0.007	0.010	0.065	0.005	0.017
	$h = 2$	2000	0.033	0.004	0.006	0.057	0.002	0.012
		4000	0.031	0.002	0.004	0.050	0.001	0.009
2	Parametric	1000	0.049	0.008	0.013	0.054	0.006	0.011
		2000	0.050	0.004	0.009	0.055	0.003	0.009
		4000	0.052	0.002	0.007	0.054	0.001	0.007
	Non-parameteric	1000	0.041	0.013	0.016	0.036	0.008	0.012
	$h = 1.1$	2000	0.036	0.007	0.010	0.033	0.004	0.007
		4000	0.031	0.004	0.006	0.030	0.002	0.005
	Non-parameteric	1000	0.054	0.008	0.014	0.071	0.005	0.020
	$h = 2$	2000	0.048	0.004	0.009	0.062	0.002	0.014
		4000	0.044	0.002	0.007	0.054	0.001	0.010
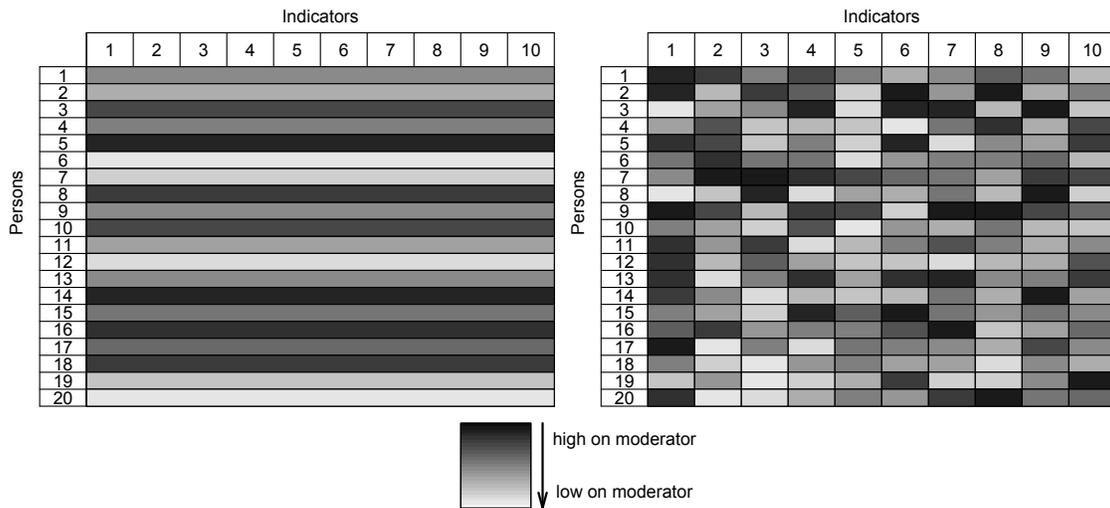
Figure 1: Indicator-level moderation modeling (on the right) as opposed to traditional moderation modeling (on the left). Traditional moderation modeling uses a single moderator which varies across persons, indicator-level moderation modeling uses indicator-specific moderators.
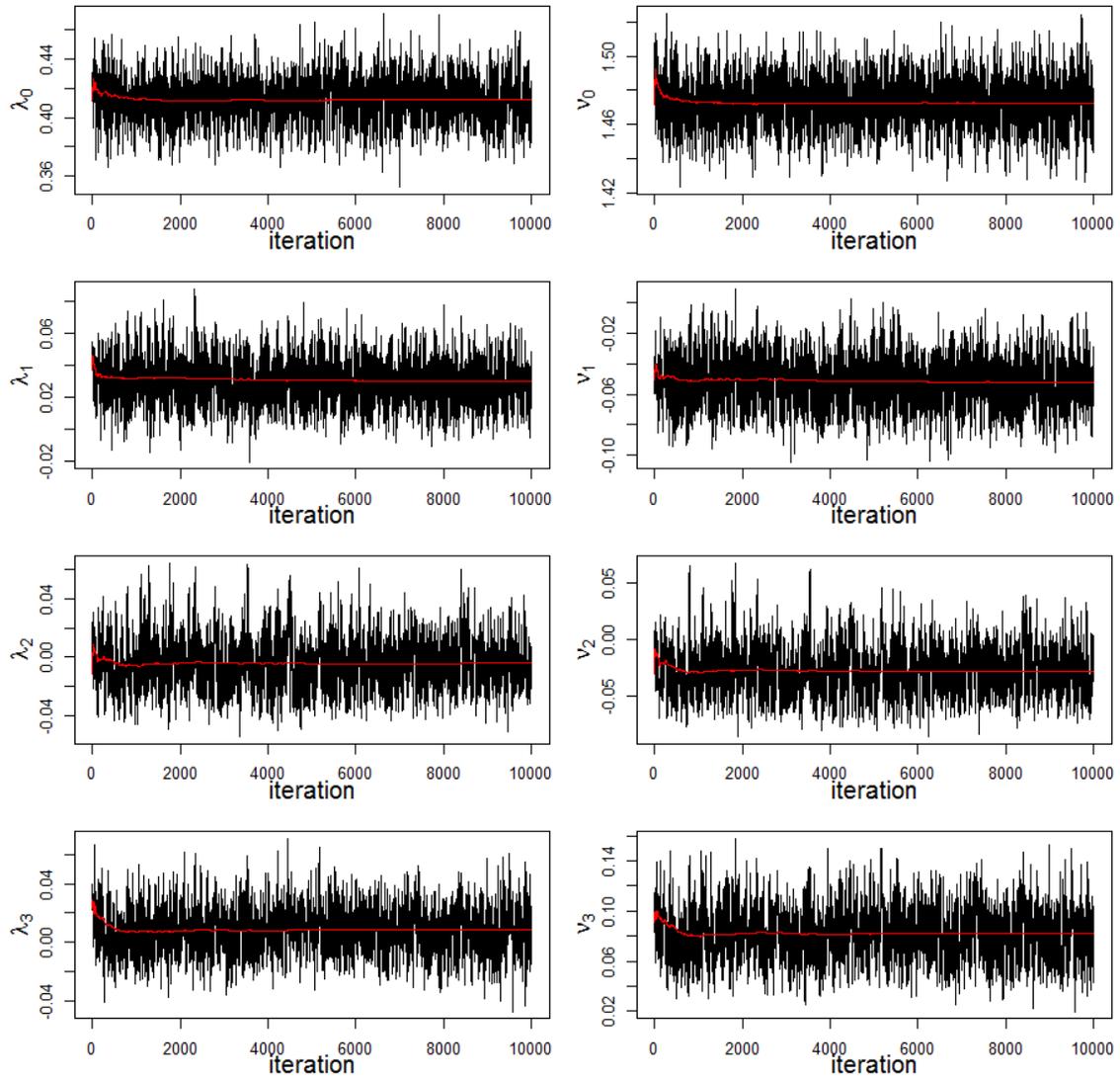
Figure 2: Example of the convergence plots (traceplots in black and running mean plots in red): parameters for item 10 in the cubic moderation model.
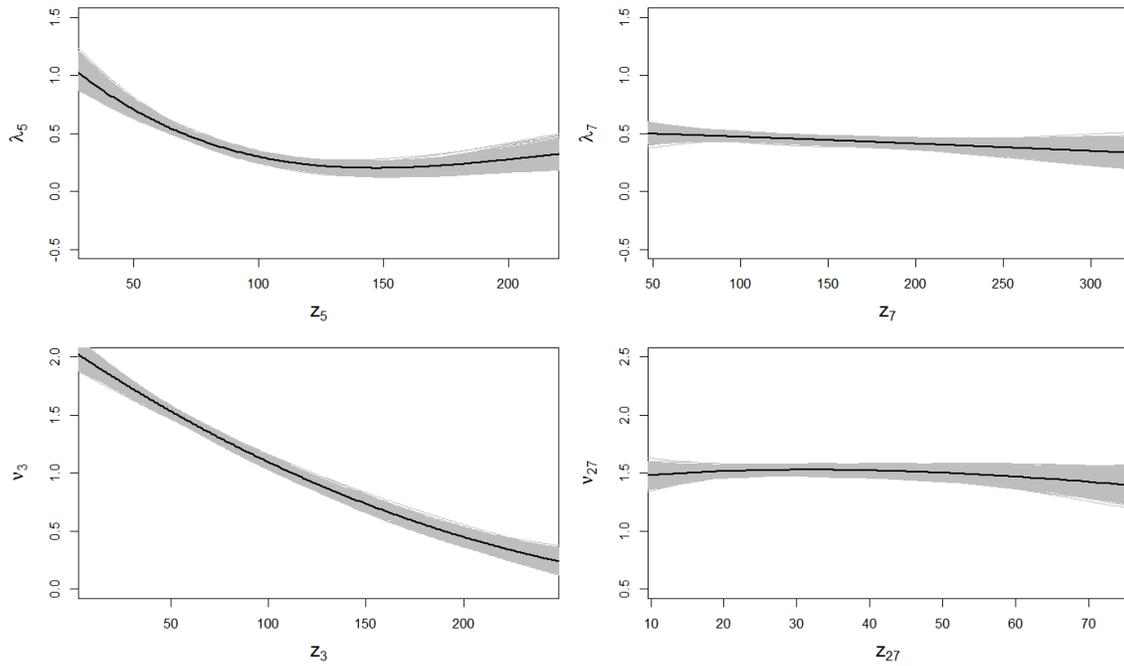
Figure 3: Example of the evaluation of the model-based uncertainty of the functional relationship between the moderator and the model parameters in the cubic moderation model. Response time is on the x-axis, and the latent variable model parameters are on the y-axis: slopes of items 5 and 7, and intercepts of items 3 and 27. Each grey line is based on a sample from the posterior distribution of the model parameters, and the black lines are based on the posterior means of the model parameters.
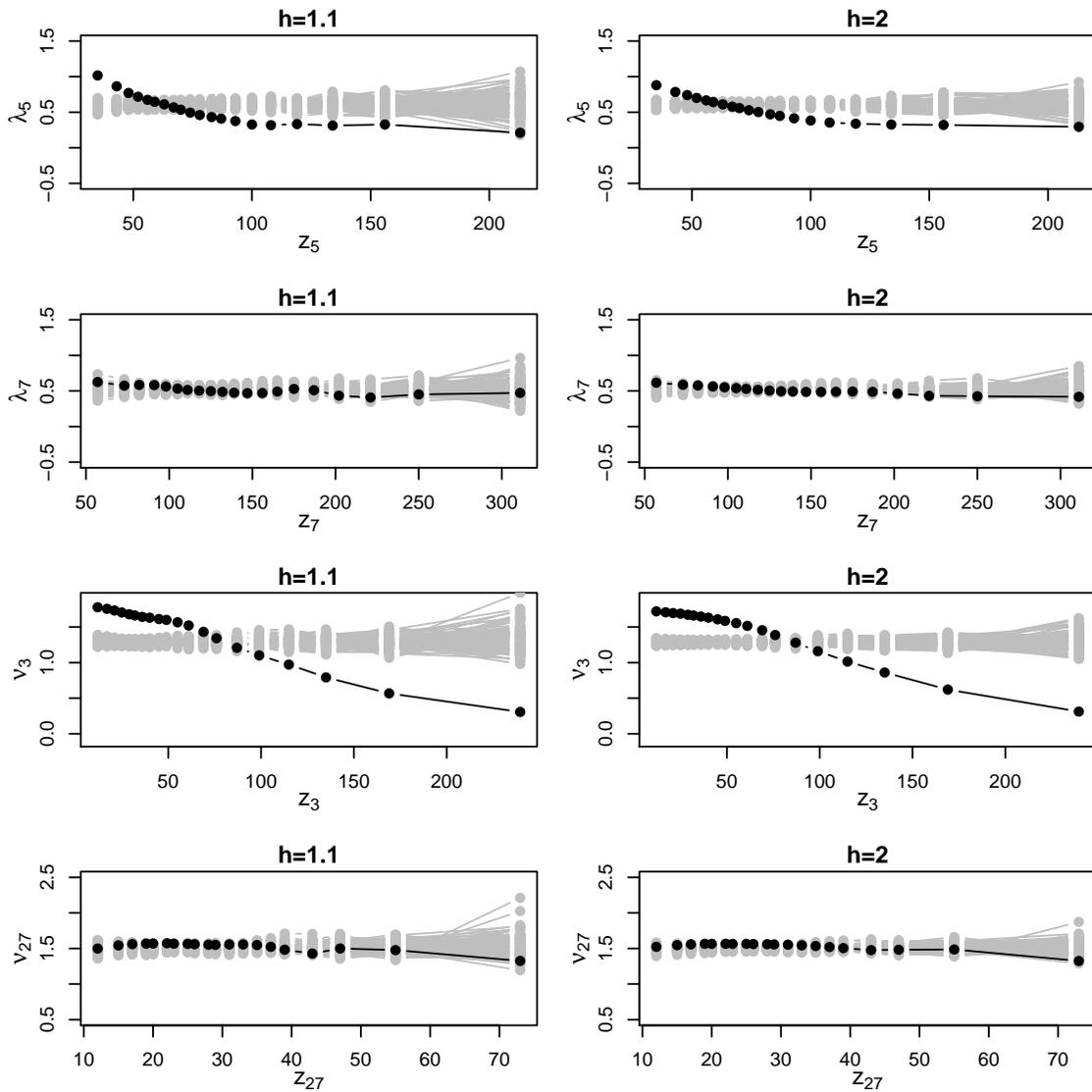
Figure 4: Example of permutation tests for the effects of the response time (on the x-axis) on the latent variable model parameters (on the y-axis): slopes of items 5 and 7, and intercepts of items 3 and 27; black lines represent the relationship estimated in the observed data, and each grey line presents the relationship estimated in each of the permuted data sets. The results are shown for the two values of the bandwidth factor ($h = 1.1$ on the left and $h = 2$ on the right)

57
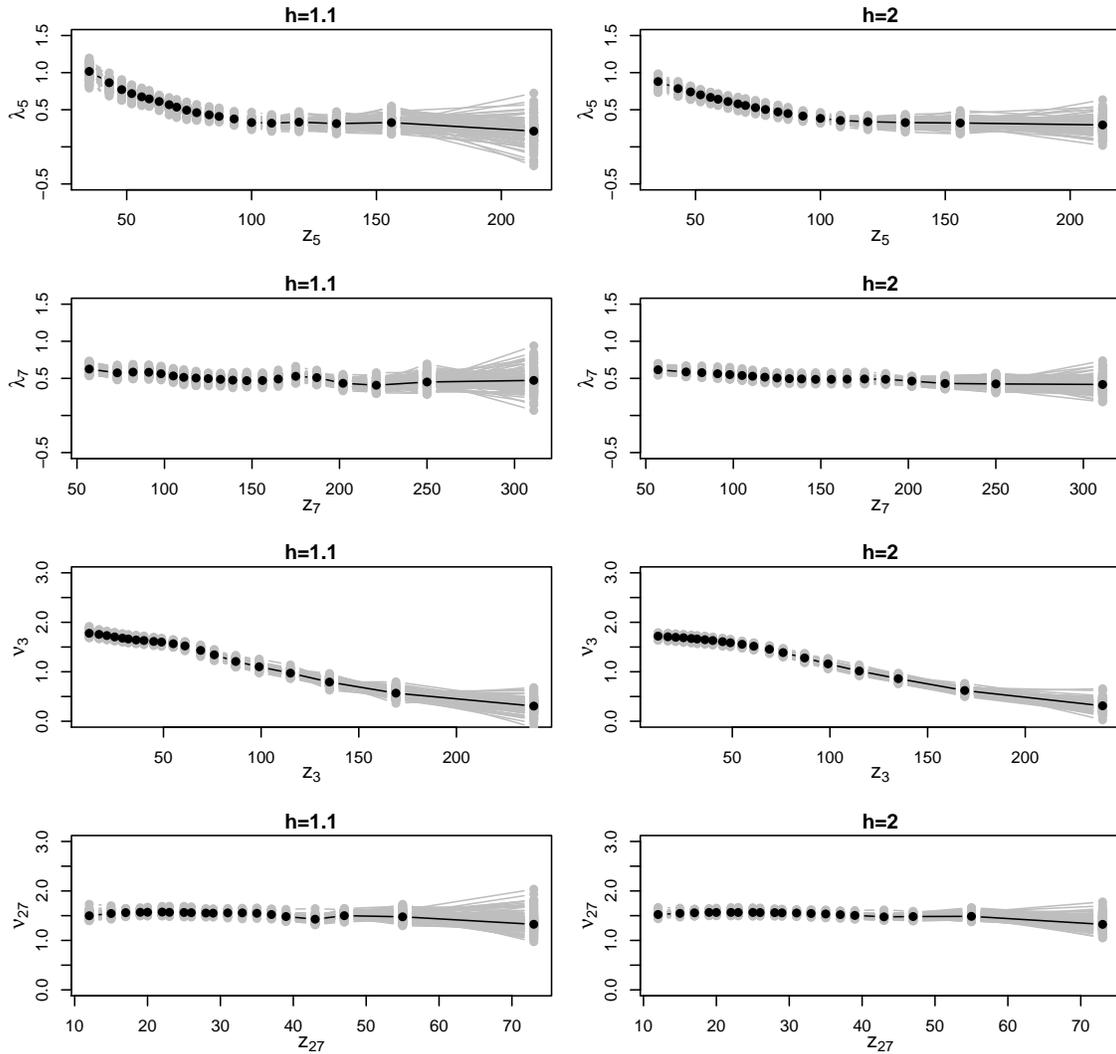
Figure 5: Example of bootstrapping for the effects of the response time (on the x-axis) on the latent variable model parameters (on the y-axis): slopes of items 5 and 7, and intercepts of items 3 and 27; black lines represent the relationship estimated in the observed data, and each grey line presents the relationship estimated in each of the bootstrap samples. The results are shown for the two values of the bandwidth factor ($h = 1.1$ on the left and $h = 2$ on the right)
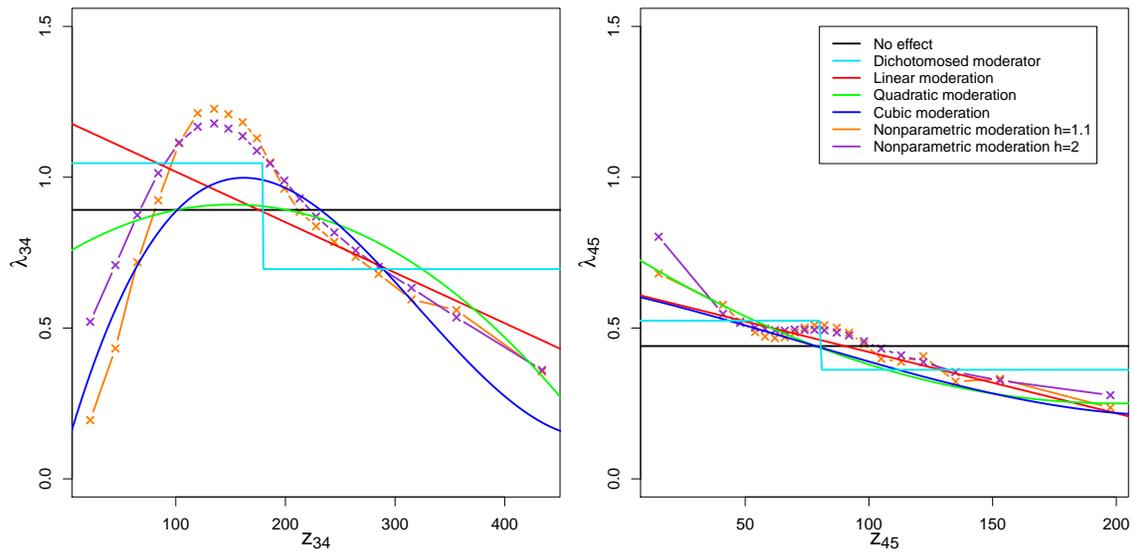
Figure 6: Example of the relationship between the moderator (z_{pi}; response time in seconds) and the slope parameter of items 34 and 45. Different lines represent different moderation methods.
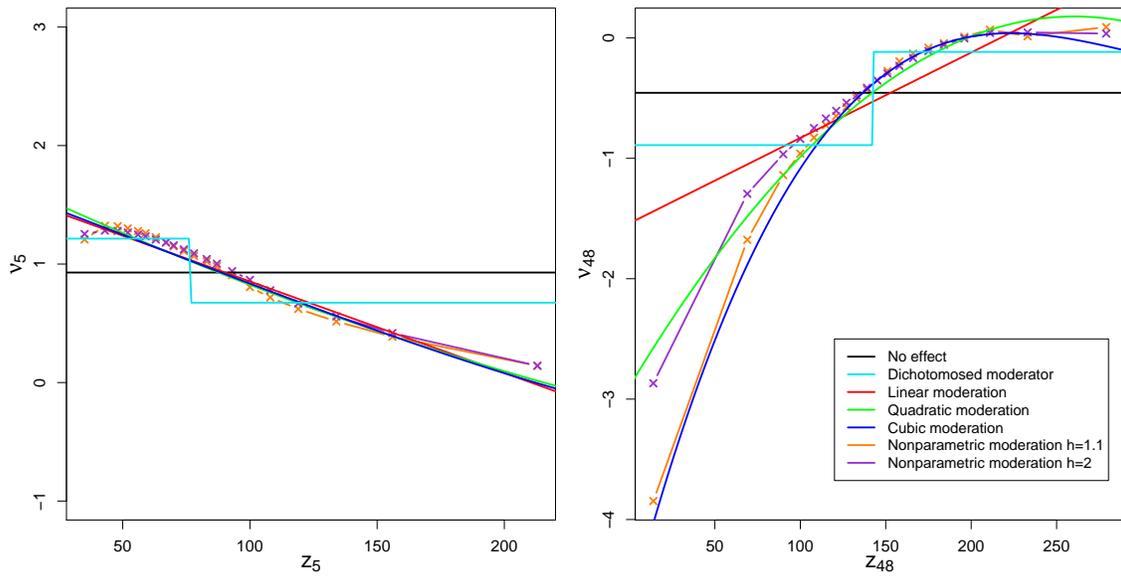
Figure 7: Example of the relationship between the moderator (z_{pi}; response time in seconds) and the intercept of items 5 and 48. Different lines represent different models.
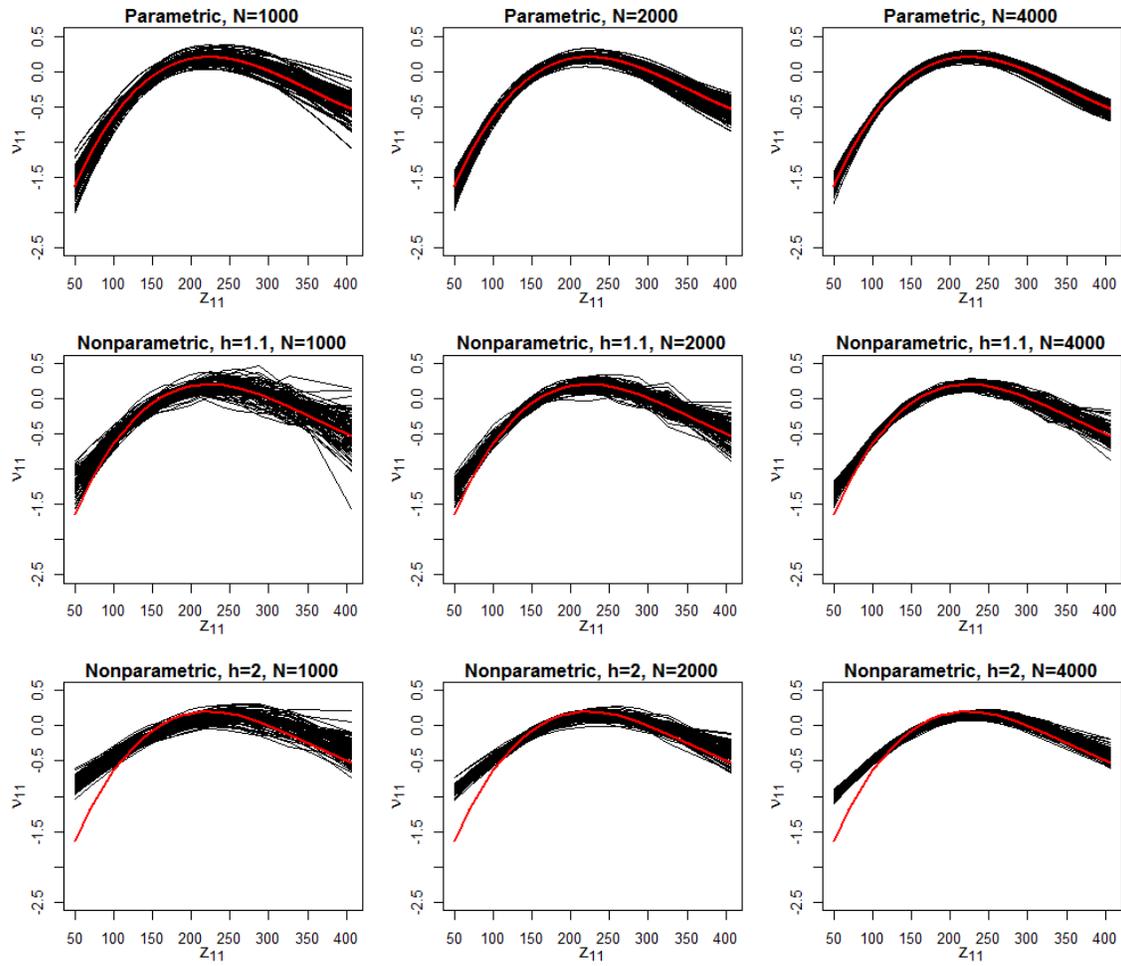
Figure 8: Recovery of the relationship between the moderator (on the x-axis) and the intercept of item 11 (on the y-axis): each black line represents the estimated relationship from one replication, the red line represents the true relationship (data were generated under the cubic model); each row represents a separate moderation method (parametric, nonparametric with the bandwidth of $h = 1.1$, and nonparametric with $h = 2$), and column represents different sample size (N).
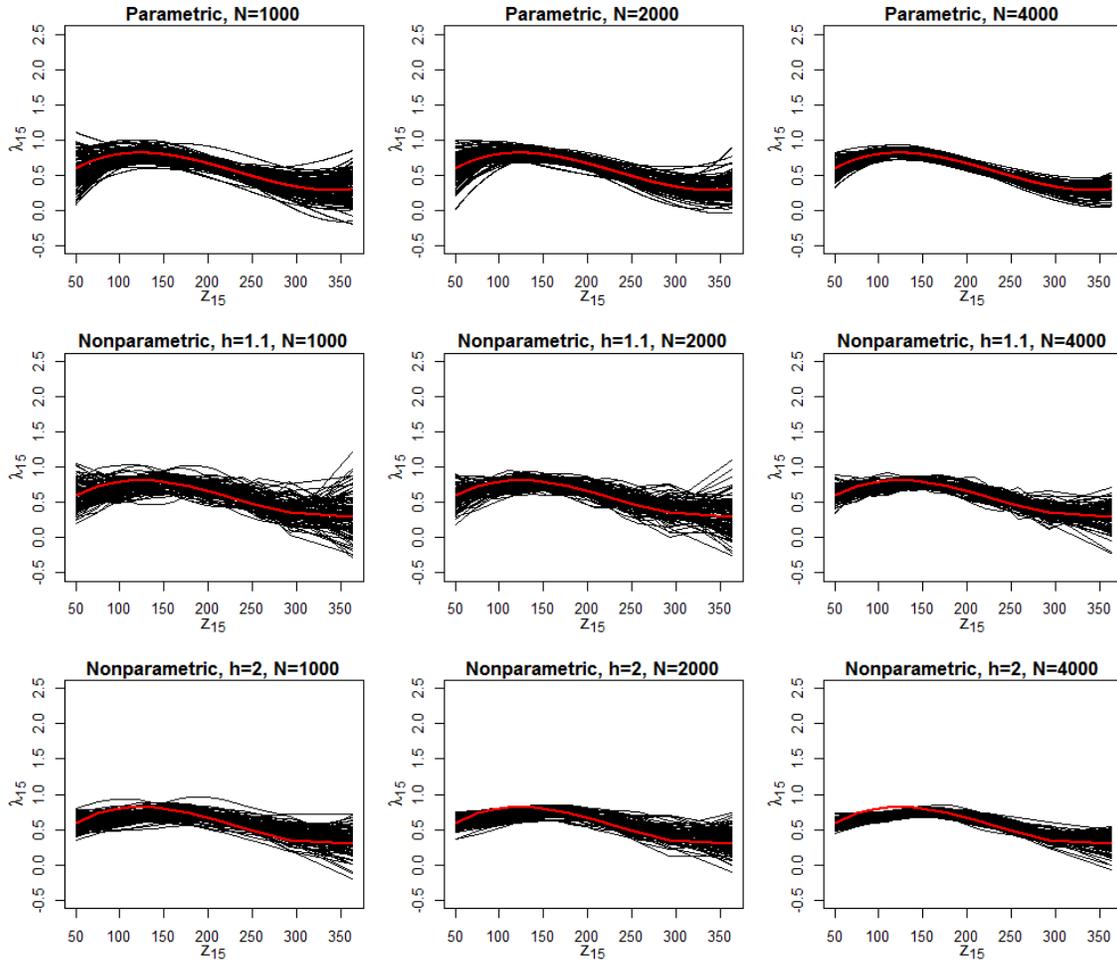
Figure 9: Recovery of the relationship between the moderator (on the x-axis) and the slope of item 15 (on the y-axis): each black line represents the estimated relationship from one replication, the red line represents the true relationship (data were generated under the cubic model); each row represents a separate moderation method (parametric, nonparametric with the bandwidth of $h = 1.1$, and nonparametric with $h = 2$), and column represents different sample size (N).
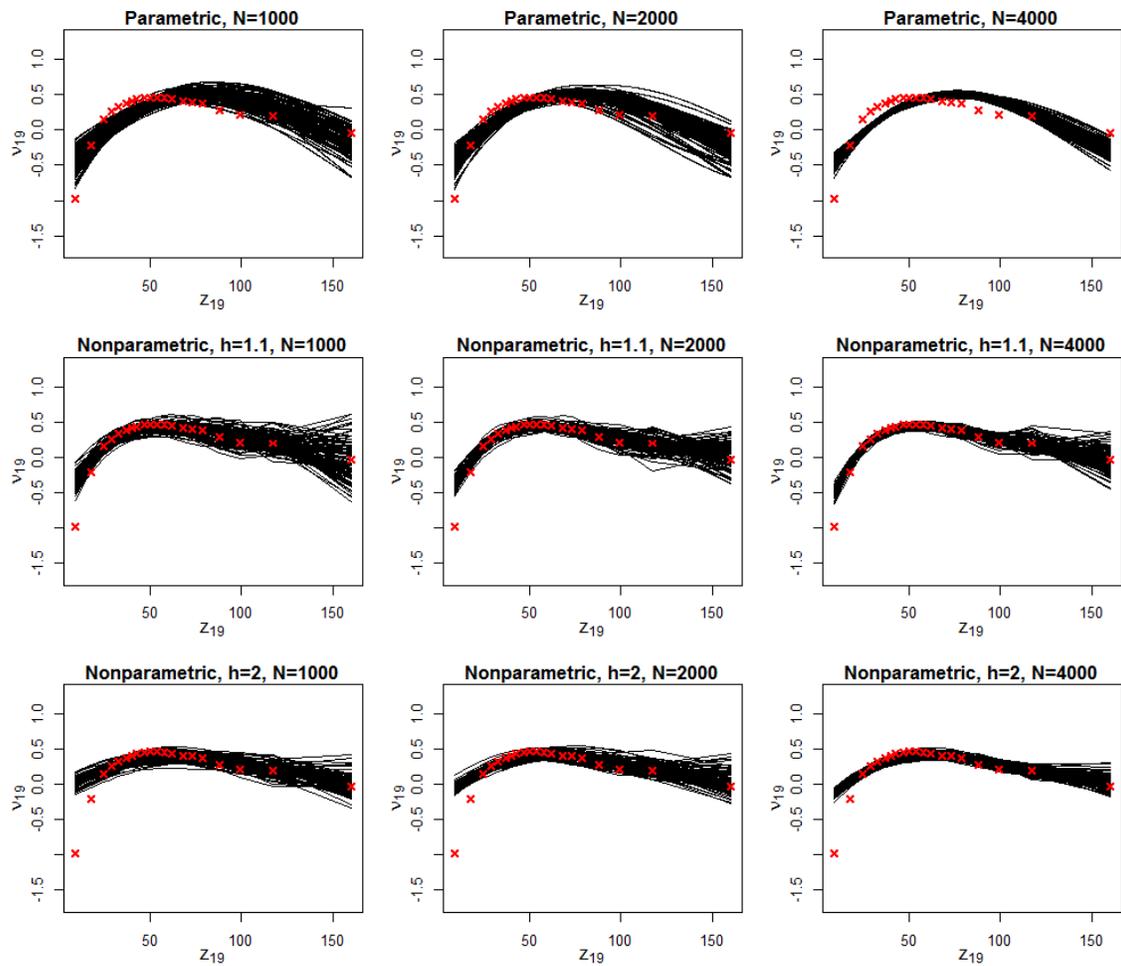
Figure 10: Recovery of the relationship between the indicator-specific moderator (on the x-axis) on the intercept of item 19 (on the y-axis): each black line represent the estimated relationship from one replication, the red dots represent the true parameters at the focal points (data were generated using the response-specific parameters obtained from the empirical data); each row represents a separate moderation method (parametric, nonparametric with the bandwidth of $h = 1.1$, and nonparametric with $h = 2$), and column represents different sample size (N).
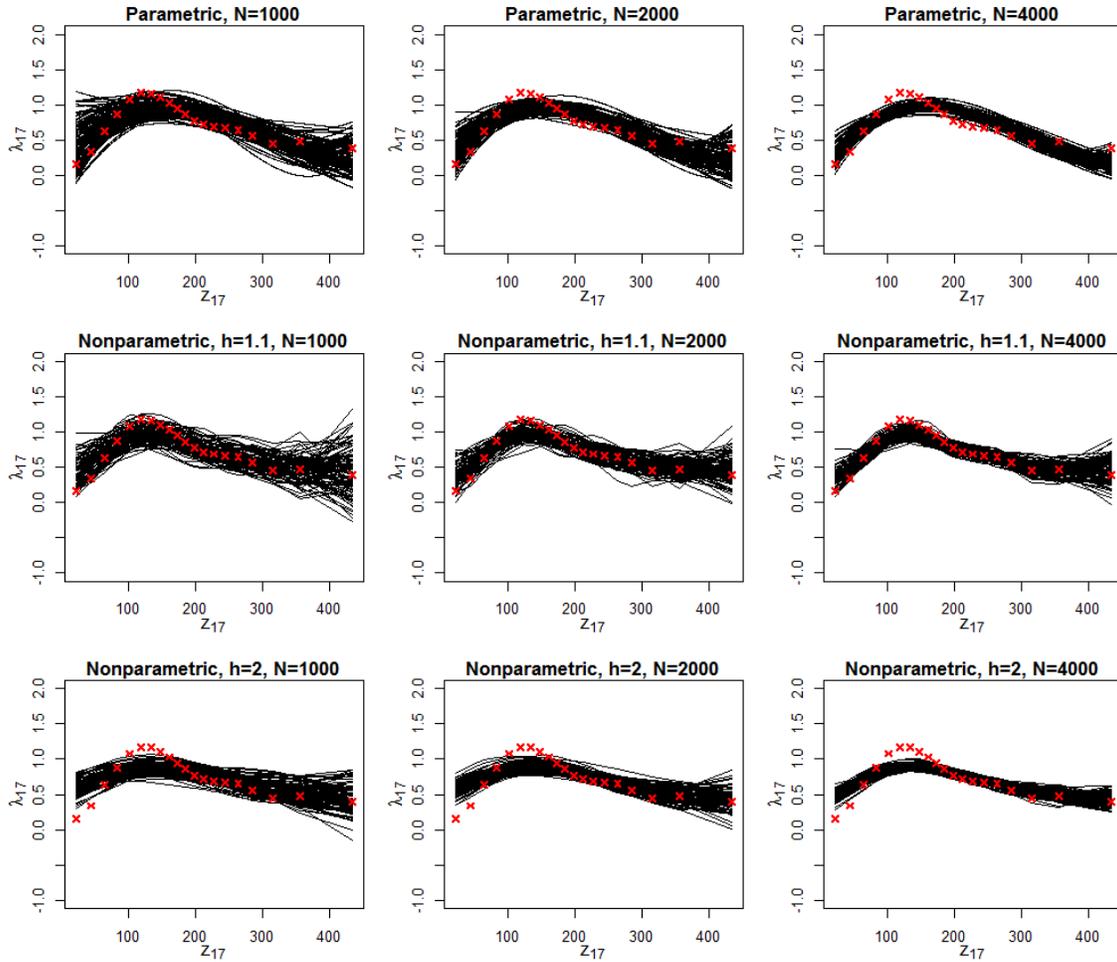
Figure 11: Recovery of the relationship between the indicator-specific moderator (on the x-axis) on the slope of item 17 (on the y-axis): each black line represent the estimated relationship from one replication, the red dots represent the true parameters at the focal points (data were generated using the response-specific parameters obtained from the empirical data); each row represents a separate moderation method (parametric, nonparametric with the bandwidth of $h = 1.1$, and nonparametric with $h = 2$), and column represents different sample size (N).