



Unimodality in personality:

Can generalized linear mixed models identify a unimodal relationship between the occurrence of anxiety and depressive disorders and the Big Five personality traits?

L.H.J. Braggaar BSc.

Master thesis Methodology and Statistics
Faculty of Social and Behaviour Sciences – Leiden University
November 2015
Student number: s1137638
Prof.dr. M.J. de Rooij

Acknowledgements

I would like to thank Mark de Rooij for his guidance and support in the last half year. As my supervisor he was always willing to point me to the right direction during the study. While I was writing this thesis, his door was always open for me to give me feedback. He was able to inspire me to overcome challenges and encouraged me to continue my research when I was on the right track.

Abstract

In psychology mostly linear analysis methods are used to analyze research data. However, in for example personality research non-linear relationships may be more likely. This study examines whether generalized linear mixed models (GLMM) can identify a unimodal relationship between the occurrence of anxiety and depressive disorders and the Big Five personality traits. Unimodal data with different number of disorders, Big Five variables and sample sizes are simulated from the Gaussian logistic curve. A GLMM is fitted to the binary data with respect to the Big Five personality factors and a statistical test tests for unimodality. An analysis of power and an analysis of bias evaluate the appropriateness of the GLMM. The GLMM turns out to be able to detect a unimodal relationship with high power and the bias values are acceptable. The GLMM is fitted to the data of the Netherlands Study of Depression and Anxiety. A statistical test points out that conscientiousness, extraversion, agreeableness and neuroticism have a unimodal relationship with the occurrence of anxiety and depressive disorders.

Table of contents

CHAPTER TITLE	PAGE
1. Introduction	
1.1 <i>Personality in psychology</i>	5
1.2 <i>Current analysis models</i>	6
1.3 <i>Unimodal relationship</i>	7
1.4 <i>Generalized linear mixed models to detect unimodality</i>	9
1.5 <i>NESDA</i>	11
1.6 <i>Unimodal personality data</i>	11
2. Methods	
2.1 <i>Monte Carlo simulation</i>	13
2.1.1 <i>Statement of the research problem</i>	13
2.1.2 <i>Experimental plan</i>	13
2.1.3 <i>Extension to five predictors</i>	14
2.1.4 <i>Simulation</i>	15
2.1.5 <i>Estimation</i>	15
2.1.6 <i>Replication</i>	16
2.1.7 <i>Analysis of output</i>	16
2.2 <i>Analysis of NESDA data</i>	17
3. Results	
3.1 <i>Results simulated data</i>	19
3.1.1 <i>Power analysis</i>	20
3.1.2 <i>Bias analysis</i>	21
3.2 <i>Results NESDA data</i>	26
4. Discussion	
4.1 <i>The simulation study</i>	30
4.2 <i>NESDA analysis</i>	31
4.3 <i>The fit of the GLMM and assumptions</i>	31
4.4 <i>Implications</i>	32
Bibliography	34
Appendix 1	36

1. Introduction

Personality is a core concept in psychology. Personality involves the characteristics that makes a person unique. It involves a great field of research, for example the relation between personality and important life outcomes and mental disorders. Linear models are mostly used to analyze these kind of research data, regardless of the certainty of linearity. This study examines whether a generalized linear mixed model (GLMM) is able to identify a unimodal relationship. Further, the relationship between the occurrence of anxiety and depressive disorders and Big Five personality traits is examined using a generalized linear mixed model.

1.1 Personality in psychology

An important field in psychology concerns personality. Personality refers to individual differences in patterns of thinking, feeling and behaving. People differ in personality characteristics and an understanding of personality allows psychologists to predict how people will respond in different situations. The five core personality dimensions are represented in the Big Five personality model. The Big Five personality model consists of openness to experience, conscientiousness, extraversion, agreeableness and neuroticism. It is well established that personality traits of patients with a mental disorders differ significantly from the traits of other persons (Cuijpers, van Straten & Donker, 2004). Research on the nature of the relationship of personality traits and mental disorders may help to clarify the etiologies of the disorders and to provide helpful support.

Roberts, Kuncel, Shiner, Caspi and Goldberg (2007) show the importance of personality. They reviewed several studies to test the predictive validity of personality traits. Traditionally, the ability of personality traits to predict important live outcomes is questioned because of small effects. Roberts and colleagues compared the effect sizes of socioeconomic status (SES) and cognitive ability to the effect sizes of personality traits. They chose three important life outcomes: mortality, divorce and occupational attainment. Based on the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism) they examined to what extend personality traits predicted those outcomes. For a fair comparison the results of the different studies needed to be transformed to a common metric. In this case, the linear association measure Pearson product-moment correlation coefficient was used. The personality traits clearly predicted important live outcomes. They even found larger effect sizes for personality traits in predicting important life outcomes than

SES and cognitive ability. People with certain personality characteristics were more likely to experience important life outcomes than others. For example traits associated with neuroticism, such as being anxious, increased the probability of divorce. Several studies showed that high levels of conscientiousness were related to a longer life (Friedman et al., 1993; Weiss & Costa, 2005; Christensen et al., 2002, as cited in Roberts et al., 2007).

Spinhoven, de Rooij, Heiser, Smit & Penninx (2009) compared the rate and pattern of comorbidity of affective disorders in relation to personality traits in patients in primary care versus specialty health care. Specialty care patients showed higher scores on the traits of neuroticism and lower scores on the traits of extraversion and conscientiousness. This may be related to the higher degree of affective disorders in specialty care patients.

Moreover, several studies show a strong association of high neuroticism and low extraversion with anxiety and depressive disorders (Jylhä and Isometsä, 2006; Bienvenu, Samuels, Costa, Reti, Eaton & Nestadt, 2004; Hayward, Taylor, Smoski, Steffens & Payne, 2013).

Spinhoven, Elzinga, van Hemert, de Rooij and Penninx (2014) examined the relationship of facets of extraversion with depression and social anxiety. Lack of positive affectivity as a facet of extraversion affects both depression and social anxiety. Even lower order facets of extraversion relate to depression and social anxiety, which can be important indications for treatment of the disorder.

The Big Five personality traits appear to be of great importance for the understanding of the occurrence of mental disorders.

1.2 Current analysis models

Different analysis methods are used to explore the relationship between the Big Five personality traits and the occurrence of mental disorders.

Roberts and colleagues (2007) used the linear association measure Pearson product-moment correlation coefficient for comparison of the results of the different studies. Spinhoven and colleagues (2009) analyzed the association of personality variables with comorbidity among anxiety and depressive disorders with second-order generalized estimating equations (GEE). The GEE allows an analysis of the dependency of correlated binary outcome data. Spinhoven and colleagues (2014) studied facets of extraversion in relation to depression and social anxiety. They used structural equation modeling (SEM) for the effects of facets of neuroticism on depression and social anxiety. SEM estimates strengths of relations between observed variables and latent variables. A special SEM method that is

often used for analysis, is the Multiple Indicators Multiple Causes model (MIMIC). A MIMIC model involves using latent variables that are predicted by observed variables. MIMIC models allow a direct association between covariates and symptoms. Moreover, explanatory IRT models can be used, including person properties. Hayward and colleagues (2013) found that high levels of neuroticism and low levels of extraversion and conscientiousness related to depression. They used linear regression models and binary logistic regression models.

As the examples above, currently linear models are mostly used for the analysis of data concerning personality and mental disorders. Linear models are easy to implement and have a straightforward interpretation (Roberts, 1986). However, in other fields than psychology, such as mathematics, physics and ecology, non-linear methods have been used successfully. Think of Newton's law of gravitation, in which the force between two masses is computed by the product of the masses divided by the squared distance. Linear methods appear to be appropriate in many situations, but more and more interest for non-linear models emerges in psychology too (Mattei, 2014). Possibly non-linear models give a better representation of the true relationship in psychological data.

For example the studies described above might benefit from a non-linear model. Roberts and colleagues (2007) found some discrepancies between the outcomes of some studies. High levels of neuroticism were associated with increase of premature mortality in several studies. In contrast, two studies reported protective effects of high levels of neuroticism. The domain of agreeableness is ambiguous too. Some studies showed a protective effect of high levels of agreeableness, while others showed that high levels of agreeableness contributed to mortality. These contradictions might appear because of characteristics of the studies, such as the use or non-use of background variables. But maybe the linear association measures that are used do not describe the data well. Maybe for higher levels of agreeableness the probability of premature mortality does in fact decrease, resulting in a single peaked relationship. In that case models that allow a unimodal relationship are more appropriate.

So, in the analysis of personality assessment the use of non-linear models should be evaluated.

1.3 Unimodal relationship

For linear analysis methods, a linear relationship is assumed. In the context of personality assessment and mental disorders, this means that the probability of occurrence of a disorder increases monotonically as subject's location on a personality trait continuum increases. So,

participants with higher levels on a personality trait have a higher probability of occurrence of a disorder (Polak, 2011). The probability curve is monotonically increasing and s-shaped (Figure 1, red line). When a monotonic relation between the two variables under study is assumed, a linear analysis model is appropriate.

Instead of a linear relationship, a unimodal relationship, assuming a single peak is plausible. Especially in the fields of attitudes, preferences and personality measurement this might be more suitable. For example an item such as “I love to go to parties every week”, measuring sociability can be disagreed from below by someone who doesn’t like parties. It can also be disagreed by someone with a high level of sociability, because one party every week is not enough (Stark, Chernyshenko, Drasgow & Williams, 2006). In Figure 1, the green line shows the typical single-peaked curve. The closer an item is located near a subject’s position, the higher the probability of endorsement. For the ideal point (subject’s position) the curve reaches his top, after this, the probability decreases. In the context of personality assessment this implicates that higher levels on a personality trait may also lead to a decrease in probability of occurrence of a disorder. The appropriateness of linear analysis methods in this case is questionable.

As mentioned before, currently linear analysis methods are mostly used. But, in the field of personality research a model, which allows a unimodal relationship may give a better representation.

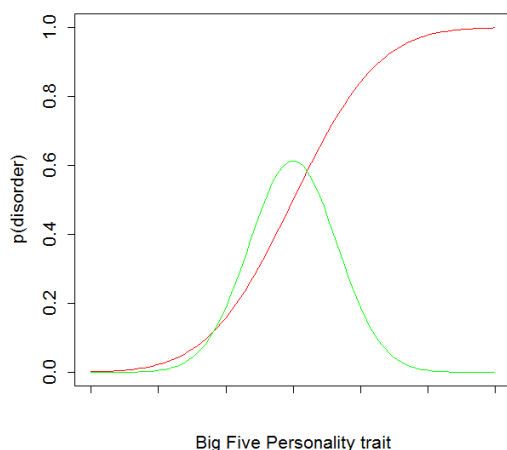


Figure 1 Example: probabilities of occurrence of a disorder to a Big Five personality trait. Red: monotonic relation, green: unimodal relation

1.4 Generalized linear mixed models to detect unimodality

In the field of ecology a promising method is used to analyze and test for unimodal data. In ecology the term niche refers to how species respond to specific environmental conditions. For example, with little predators they will grow and with little nutrients they will die. Niche theory assumes unimodal curves for species` occurrence with respect to environmental gradients. Nevertheless, usually, when applying GLM, linear (straight-line) models are fitted. Often the need to fit a model, that allows a unimodal relationship, is not recognized (Austin, 2007; Jamil & Ter Braak, 2013).

Jamil and Ter Braak (2013) use generalized linear mixed models (GLMM) to detect and analyze unimodal species-environmental relationships. They show that a GLMM can be seen as a Gaussian logistic model. Jamil and Ter Braak (2013) propose a statistical test to test for a unimodal species-environmental relationship while just fitting a generalized linear mixed model. In the following a detailed explanation of their method.

Jamil and Ter Braak (2013) started with a logistic linear mixed model with probability of occurrence p_{ij} of species j in site i to an environmental variable x_i

$$\text{logit}(p_{ij}) = \alpha_j + \beta_j x_i + \gamma_i \quad (1)$$

with slope α_j , intercept β_j and site effect γ_i all as random parameters. The random site parameter γ_i may account for factors that influence the probability of occurrence of all species in a site, such as size of the site.

Jamil and Ter Braak (2013) show that the logistic linear mixed model represented in equation (1) can in fact be interpreted as a Gaussian logistic model.

The unimodal Gaussian logistic curve is one of the simplest unimodal curves for presence-absence data (Oksanen et al., 2001; Ter Braak & Looman, 1986, as cited in Jamil & Ter Braak, 2013).

$$\text{logit}(p_{ij}) = a_j - \frac{(x_i - u_j)^2}{2t_j^2} \quad (2)$$

In the unimodal Gaussian logistic curve of equation (2) u_j corresponds to the environmental value for which the presence of species is most likely, x_i is the environmental variable per site, a_j is a coefficient related to the maximum probability of occurrence and tolerance, t_j refers to the standard deviation of the unimodal curve.

The unimodal Gaussian logistic curve (2) can be written like the logistic linear mixed model (1) by expanding the quadratic term. To avoid that γ_i also depends on j , Jamil and Ter Braak (2013) assume equal tolerances for the disorders $t_j = t$.

$$\alpha_j = a_j - \frac{1}{2t^2} u_j^2 \quad \beta_j = \frac{u_j}{t^2} \quad \gamma_i = \frac{-1}{2t^2} x_i^2 \quad (3)$$

$$\text{logit}(p_{ij}) = \alpha_j + \beta_j x_i + \gamma_i = a_j - \frac{1}{2t^2} u_j^2 + \frac{u_j}{t^2} x_i + \frac{-1}{2t^2} x_i^2$$

So, the GLMM in equation (1) can be seen as a Gaussian logistic model with equal tolerances.

Jamil and Ter Braak (2013) fit the GLMM from equation (1) to binary data of the presence or absence of species in a site with respect to the environmental variable. They propose a statistical test to test for a unimodal response. The null model is the GLMM from equation (1), with independent and normally distributed site effects. In the alternative model the site effects depend quadratically on the environmental variable. The site effects also depend on the total number of species occurring in a site (S). If the quadratic part in the alternative model, x^2 , is judged as significant in an ANOVA test on the regression coefficients of the models, this gives evidence for a unimodal response. Graphically, there is an indication for a unimodal response if the graph of the random site effects and the environmental variable shows an N-shaped relationship. The null model and alternative model are represented by

$$H_0: \gamma \sim x + S \quad H_a: \gamma \sim x + x^2 + S \quad (4)$$

Jamil and Ter Braak (2013) illustrate their method with simulated data and with data from three real data sets. They simulate data sets with different number of species (10, 50, 100) and tolerances (0.5, 1, 4), all showing a unimodal relationship, to examine whether the generalized linear mixed model is able to detect the unimodal relationship. In all data sets the squared term in equation (4) was significant.

They also illustrate their method with three real data sets. All of them describing vegetation in different sites. In each data set one environmental variable, like temperature, is examined. The number of sites in the data sets varies from 20 to 200. In each site different species, varying from 30 to 60, occurs. For each data set Jamil and Ter Braak (2013) fit the generalized linear mixed model and test for a unimodal response. In all data sets there is a

unimodal relationship. Jamil and Ter Braak conclude that a GLMM with random site effects can effectively deal with a unimodal response in the multi-species data sets.

1.5 NESDA

The Netherlands Study of Depression and Anxiety (NESDA) determines factors that influence the development and prognosis of anxiety and depression. Almost 3000 persons, with and without symptoms, are monitored for a long period of time. Among other things, validated measurements are used to assess depression and anxiety levels. Moreover, personality characteristics are examined. Compared with the ecological data, the NESDA data have five different disorders, corresponding to different species. Moreover, personality data from almost thousands of persons is available, instead of hundreds of sites in the ecological data. Finally, in the NESDA data the relationship between the occurrence of disorders is examined considering five personality disorders instead of one environmental variable.

1.6 Unimodal personality data

Jamil and Ter Braak (2013) showed in their study that generalized linear mixed models can identify a unimodal relationship in ecological data. In this study generalized linear mixed models are evaluated in the field of personality assessment. We hypothesize that many outcome variables in psychology have a unimodal relationship with personality variables. The central question in this study is whether generalized linear mixed models can identify unimodal relationships in data with characteristics unlike the data used by Jamil and Ter Braak (2013). Specifically, generalized linear mixed models are evaluated in the context of the NESDA data. A simulation study is used to test whether a generalized linear mixed model is able to identify a unimodal relationship in data with NESDA characteristics. In the simulation study the fit of the GLMM is evaluated by comparing the simulated (true) values for a unimodal curve with the by the model estimated values. If there is high correspondence, there is low bias and the model can estimate the parameters adequately. Besides bias, the root-mean-square-error (RMSE) is calculated. The RMSE is a measure of the differences between true values and estimated values. In general, the smaller the RMSE value, the better the prediction. Moreover, a statistical test tests for unimodality.

Secondly, the unimodality test and analysis are executed on the NESDA data. The test tests whether a unimodal response between the occurrence of anxiety and depressive disorders and the Big Five personality traits is present. Nowadays often linear models are used to analyze

NESDA data. However, a unimodal relation in the context of personality assessment seems reasonable. In that case, analysis methods that allow a unimodal relationship should fit the data well.

2. Methods

To examine whether generalized linear mixed models can identify unimodal relationships in data with characteristics unlike those of Jamil and Ter Braak (2013), unimodal data are simulated. Moreover, the GLMM can be fitted to the NESDA data.

2.1 Monte Carlo simulation

Data are simulated to evaluate the appropriateness of generalized linear mixed models for data with characteristics unlike the data used by Jamil and Ter Braak (2013). The six-step approach for Monte Carlo simulation designs recommended by Skrondal (2000) and Paxton, Curran, Bollen, Kirby and Chen (2001) is used to present the design of the simulation study.

2.1.1 Statement of the research problem

This study examines the appropriateness of generalized linear mixed models for the analysis of possible unimodal data in psychological context. As discussed before, Jamil and Ter Braak (2013) conclude that generalized linear mixed models can detect a unimodal relationship between the occurrence of species in different sites considering an environmental variable. The generalized linear mixed model that Jamil and Ter Braak fitted to their data is appropriate for the characteristics of the ecological data. In order to test the appropriateness of the model for psychological data, unimodal response data are simulated with characteristics of the NESDA data. In this case the logistic linear mixed model relates the probability of occurrence p_{ij} of a DSM disorder j in person i to a quantitative Big Five personality variable x_i . The random person effects γ_i may account for factors that influence the probability of occurrence of all disorders in a person, such as family support. This simulation study explores the influence of sample size, number of Big Five personality factors and number of DSM disorders on the appropriateness of generalized linear mixed models.

2.1.2 Experimental Plan

Only a limited number of conditions can be investigated in a Monte Carlo experiment. The appropriateness of generalized linear mixed models is evaluated for different combinations of the factors mentioned above. In psychological research often large sample size is preferred; in the NESDA data even 3000 persons participated the study. In this simulation study the fit of GLMM is examined for sample size 50 and for larger sample sizes. Moreover, in this

simulation study we extend up to five Big Five personality factors. Finally, fewer DSM disorders are used, corresponding to species. This results in a 3x3x2 design, represented in Table 1.

	Levels	Ecological data
Sample size (# persons)	50	20-200 sites
	400	
	600	
Big Five personality	1	1 environmental variable
	2	
	5	
DSM disorders	5	10-100 species
	10	

Table 1 Variables considered in the Monte Carlo study

2.1.3 Extension to five predictor variables

In this simulation study we extend up to five Big Five personality factors. In equation (3) is illustrated how the unimodal model with one personality variable can be rewritten as a linear mixed model. The unimodal model with more than one personality variable can similarly be rewritten as a linear mixed model. See Appendix 1 for a detailed extension to three and more variables. No higher order interaction effects appear; only pairwise interaction effects. In this simulation study we assume the variances of the predictors to be equal ($d_1 = d_2 = d_3 = d_4 = d_5$). Moreover, we assume the absence of covariances between the predictors ($d_{12j}, d_{13j} \dots = 0$). d 's are precision parameters and correspond to $1/t$.

In the simulation study data are simulated with constant tolerance ($t = 1$), relating to the standard deviation of the unimodal curve. In the Gaussian logistic model with more than one Big Five personality factor the subtraction of the disorders' precision parameters (d_j) leads to lower probabilities. With five personality factors, this eventually results in constant binary y data. To overcome this problem, d is simulated as .5 in the situation with two personality factors and as .25 in the situation with five personality factors. In those cases the produced probabilities are similar to the probabilities with one personality factor. Jamil and Ter Braak (2013) showed in their simulation study that a GLMM with a tolerance of 4 (corresponding to $d = .25$) can detect a unimodal relationship.

2.1.4 Simulation

In this simulation study we simulate n values for x_{ik} variables corresponding to the Big Five personality traits, with k representing the number of the Big Five variable and i representing person. We simulate data with constant tolerance ($t = 1$) and for m disorders. Binomial probabilities are computed from the unimodal curve for presence-absence data, the Gaussian logistic curve. Finally, binary data, y_{ij} , are generated from a Bernoulli distribution.

The procedure to simulate data is the following:

- (1) Generate n values of the Big Five personality traits ($x_{i1} - x_{ik}$) as a random sample from a normal distribution, $x \sim N(0, 1)$.
- (2) Generate k vectors u (u_{kj}) of length m from a uniform distribution $U(-s, s)$, where $s = 2+c$, for a fixed value of c .
- (3) Generate a vector a of length m as a random sample from the standard normal distribution.
- (4) Compute binomial probabilities p_{ij} from the unimodal response curve.

With one Big Five variable:

$$p_{ij} = \text{logit}^{-1} \left(a_j - \frac{(x_i - u_j)^2}{2t_j^2} \right)$$

With two Big Five variables:

$$p_{ij} = \text{logit}^{-1} \left(a_j - \frac{1}{2} (d_1(x_{i1} - u_{1j})^2 + d_2(x_{i2} - u_{2j})^2) \right)$$

With five Big Five variables:

$$p_{ij} = \text{logit}^{-1} \left(a_j - \frac{1}{2} (d_1(x_{i1} - u_{1j})^2 + d_2(x_{i2} - u_{2j})^2 + d_3(x_{i3} - u_{3j})^2 + d_4(x_{i4} - u_{4j})^2 + d_5(x_{i5} - u_{5j})^2) \right)$$

- (5) Generate binary data y_{ij} at random from a Bernoulli distribution.

2.1.5 Estimation

In the R package lme4 a generalized linear mixed model (1) is fitted to the simulated binary data to test for a unimodal relationship.

`lmer(y ~ 1 + x + (1 + x | disorder) + (1 | person), family = binomial(link = "logit"), data)`

y represents the binary response data. Disorder and person represent the random slope and person effects.

Appendix 1 shows how generalized linear mixed models with more than one personality factor can be fitted using the package lme4. At this point we assume the covariances to be absent and variances of the predictors to be equal.

2.1.6 Replication

With more experimental conditions the results can better be generalized beyond the experiment. A large number of replications in a Monte Carlo experiment improves precision. But, due to limitations of computer resources often a researcher must make a trade-off between external validity and precision (Skrondal, 2000). In this study steps 3 and 4 are replicated 100 times.

2.1.7 Analysis of output

The statistical test represented in equation (4) tests whether there is a unimodal relationship between the Big Five personality traits and the occurrence of a disorder. In this case the site effects correspond to random person effects. Because the data are simulated as unimodal, this test is expected to be significant. For each condition the significance test is executed 100 times. The power of the test is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true. This is computed by dividing the number of significant tests by the total number of replications.

The GLMM estimates values for α_j , β_j and γ_i . Based on those values we can calculate values for a_j , u_{kj} and t/d using equations (3) for one Big Five variable and the equations described in Appendix 1 for more Big Five variables. It is possible that GLMM estimates the random person effects (γ_i) as positive which makes the “back” calculation (equations 3) impossible, because a root must be extracted from $-x_i^2$ divided by $2\gamma_i$. For that reason a constant, c , is subtracted from γ_i and added to α_j .

The calculations give us i number of tolerance estimates. We assumed equal tolerances for the disorders, so we take the median of the estimates to get the best estimate for t .

The estimated values can be compared to the true values for a_j , u_{kj} and t/d to get an indication of the fit of the model. If the GLMM estimates show low bias, the model is appropriate for data with NESDA characteristics. Bias can be calculated as following with $\hat{\Theta} / \Theta$ referring to the (estimated) values a_j , u_{kj} and t/d and n to the number of replications

$$\text{Bias} = \frac{\sum(\hat{\Theta} - \Theta)}{n} \quad (5)$$

The Root-Mean-Square-Error (RMSE) measures the differences between values estimated by the model and the true values. The RMSE summarizes the magnitudes of the errors in a single measure. The RMSE is calculated as following

$$\text{RMSE} = \sqrt{\frac{\sum(\hat{\theta} - \theta)^2}{n}} \quad (6)$$

The RMSE values can be compared to the RMSE value of a data set with characteristics of Jamil and Ter Braak (2013). RMSE is the root of the mean-square-error (MSE). MSE is the variance of the estimates plus the squared bias. So by calculating bias and RMSE we can draw inferences about the spread of the estimates.

2.2 Analysis of NESDA data

If the GLMM turns out to describe the data with NESDA characteristics well, a GLMM is fitted to the NESDA data. Data are from an 8-year longitudinal cohort study, the Netherlands Study of Depression and Anxiety (NESDA). Since 2004 up to 3000 people participated the study. People with and without symptoms of all ages are monitored for a long period of time. NESDA recruited respondents from the general population, primary care and mental health organizations.

Depressive disorders under study are major depressive disorder (MDD) and dysthymia (Dys). Anxiety disorders under study are social phobia (SP), panic disorder (PD), agoraphobia (AGO) and generalized anxiety disorder (GAD). Besides depression and anxiety symptoms, biological and genetic factors are examined.

Data are collected using face-to-face interviews, written questionnaires and medical examination. The presence of anxiety and depressive disorders was assessed using the depression and anxiety sections of the Composite International Diagnostic Interview (CIDI). Results are presented dichotomous with 1 referring to the presence of the disorder and 0 the absence of the disorder.

Personality was assessed using the NEO-five factor inventory (NEO-FFI), measuring openness, conscientiousness, extraversion, agreeableness and neuroticism. An exclusion criteria was the presence of a psychiatric diagnosis other than depression or anxiety.

The main aim of NESDA is to determine the factors that influence the development and prognosis of depression and anxiety (Penninx, Beekman, Smit, Zitman, Nolen, Spinhoven, Cuijpers, de Jong, van Marwijk, Assendelft, van der Meer, Verhaak, Wensing, de

Graaf, Hoogendijk, Ormel & van Dijck, 2008). In this study all disorders are used except agoraphobia. In total data from 600 respondents is available.

Generalized linear mixed models with one and five Big Five personality factors are fitted to the NESDA data and a unimodal test is executed to examine the relationship between the occurrence of anxiety and depressive disorders and the Big Five personality traits. To make inferences about the direction of the relationship, the values for u_j , a_j and t_j are computed. Jamil, Kruk and Ter Braak (2014) rewrote the Gaussian logistic model (2) as a generalized linear model defined as a second degree polynomial (Equation 7). For convenience they dropped the indices i and j .

$$\text{logit}(p) = b_0 + b_1x + b_2x^2 \quad (7)$$

By fitting this model as a generalized linear model, the estimates of the Gaussian parameters can be found the following

$$u_j = -\frac{b_1}{2b_2} \quad t_j = \sqrt{-\frac{1}{2b_2}} \quad a_j = b_0 - \frac{b_1^2}{4b_2} \quad (8)$$

3. Results

First the results of the simulation study are discussed. Next the results of the GLMM applied to the NESDA data are presented.

3.1 Results simulated data

Data sets with varying characteristics, described in the experimental plan, are simulated to evaluate the appropriateness of the generalized linear mixed model in identifying a unimodal relationship. Figure 2 shows the simulated unimodal response (occurrence probability p_{ij} for a person having a disorder) against one Big Five variable x . Graphs for different sample sizes and number of depression and anxiety disorders are presented. In this example tolerance is set as 1.

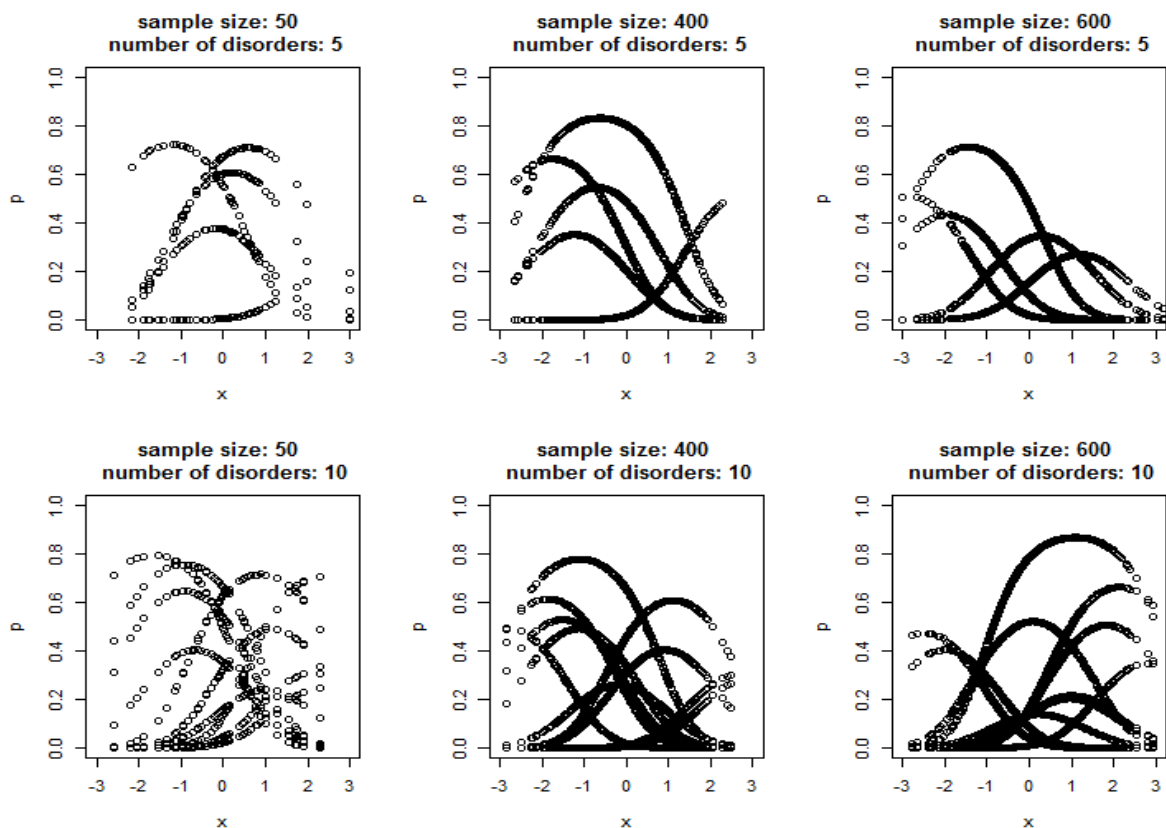


Figure 2 Simulated unimodal response against a single Big Five variable

A generalized linear mixed model (1) is fitted to the simulated data to test for a unimodal relationship. Graphically, there is an indication for a unimodal relationship if the plot of the random person effects (γ_i) and Big Five variable shows an N-shaped quadratic relationship. Figure 3 shows the graphs for different sample sizes and number of disorders for one Big Five variable. In all graphs the N- shape is recognizable.

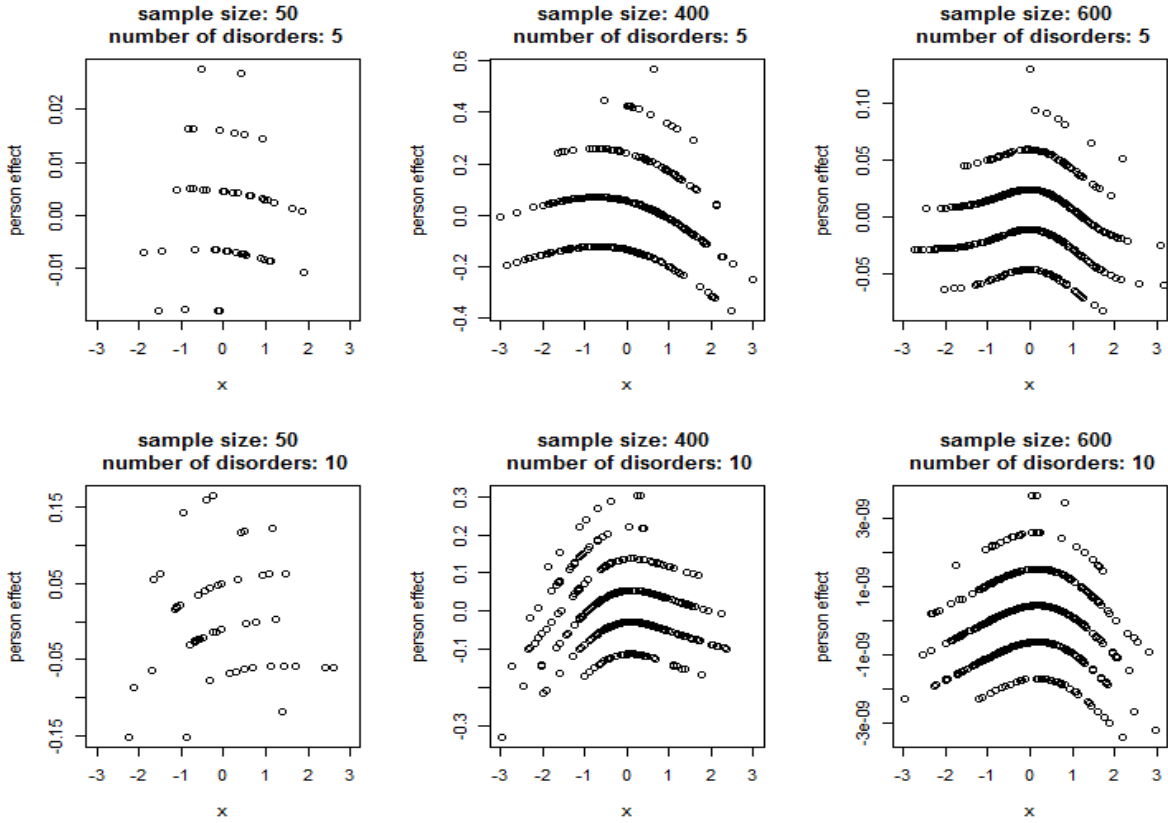


Figure 3 Estimated person effects GLMM against a single Big Five variable

3.1.1 Power analysis

The statistical test in equation (4) for a unimodal response is executed to all data sets. If the generalized linear mixed model detects a unimodal relationship, the null hypothesis is rejected ($\alpha = .05$). For each condition the test is replicated 100 times. The power of the test is the proportion of times the null hypothesis is rejected. The higher the power of the test, the less likely to make a type II error (concluding there is no effect, when in fact there is). In Table 2 the power indices are shown.

		# DSM disorders	
		5	10
Big Five variables	Sample size		
1	50	.94	.99
	400	1.0	1.0
	600	.98	1.0
2	50	1.0	.97
	400	.99	1.0
	600	1.0	1.0
5	50	.92	.98
	400	1.0	1.0
	600	1.0	1.0

Table 2 Power unimodal test

In conditions with the smallest sample size the power indices are slightly smaller. But in all conditions the power index is higher than .8, indicating high power. The GLMM seems to be able to detect unimodal relationships in our data set.

3.1.2 Bias analysis

Besides the analysis of power of the generalized linear mixed model, a bias analysis is performed. For each data set the GLMM parameter estimates are compared with the simulated (true) parameters. The GLMM gives parameter estimates for α_j , β_j and γ_i . Making use of the equations in equation (3), estimates for a_j , u_{kj} and t/d are computed and compared to the true values.

The estimated values for a_j , u_{kj} and t/d are compared with the true values and the bias and RMSE are calculated. Table 3 presents the bias values. Red values indicate negative bias, in other words, the estimated values are smaller than the true values. The left part of Table 3 indicates the condition. The first column indicates the number of Big Five variables, the second column indicates the sample size and the third column indicates the number of disorders. Horizontal the estimated values, a_j , u_{kj} and t/d , are presented. For more than one Big Five variable only the first set of u_{kj} bias values are shown; the other sets are comparable.

The ideal situation is that all parameters are estimated perfectly and there is no bias. This is the case in neither of the conditions. We start with the condition with one Big Five variable. This is the condition that corresponds mostly to the conditions of Jamil and Ter Braak (2013). The bias for a_j , the maximum probability of occurrence, is relatively high. Most bias values are above one. Remarkable too is that the a_j bias values are always positive. Apparently the estimated values are larger than the true values. The estimates for u_{kj} show much less bias. The majority of the bias values is around .1. The bias values for five and ten disorders do not differ from each other. Also sample size doesn't seem to influence bias values; bias for sample size 50, 400 and 600 are roughly equal.

In the conditions with two and five Big Five personality variables, bias for a_j is still clearly larger. Overall, for more Big Five personality variables the bias for a_j becomes larger. The estimates for u_{kj} represent the x -value for which the disorders' curve reaches his optimum for each predictor. It equals the bias for one Big Five personality variable. Again the use of five or ten disorders doesn't seem to have big impact on the bias. With respect to sample size there aren't big differences between $n = 50$, $n = 400$ and $n = 600$. Only in the situation with five Big Five variables the bias with sample size 400 is slightly smaller than with sample size 50 and in the situation with two Big Five variables bias for d is clearly smaller for sample sizes 400 and 600.

Table 4 presents the RMSE values. RMSE values summarize the magnitudes of the errors in a single measure. The lower the RMSE value, the better the fit. For one Big Five variable the RSME values for t , u_{kj} , a_j are roughly equal in all conditions. The RMSE values for two and five Big Five variables are somewhat smaller, especially for u_{kj} estimates. An increase in sample size often leads to smaller RMSE values, especially from 50 to 400 in the situation with five Big Five personality factors and for t and u_{kj} estimates in situation with two Big Five variables.

Figure 4 displays some RMSE values visually. The upper left figure displays the RMSE values for different sample sizes in the condition of five disorders and a single Big Five variable. The RMSE values for t are clearly smaller than for the other parameters. Moreover, there is few difference between u_{kj} and a_j RMSE values and for different sample sizes. The upper right figure displays the RMSE values for two Big Five variables. All values are somewhat smaller than for a single Big Five variable, especially RMSE values for u_{kj} estimates. Moreover, RMSE values for sample size 50 are slightly higher than for sample size 400 and 600. The bottom figure (for five Big Five variables) illustrates this even better.

By combining the RMSE values and bias values we can draw inferences about the standard deviation of the estimates. The RMSE incorporates the root of the squared bias plus the variance of the estimates. By visual inspection of the Table 3 and 4 we can conclude that the standard deviation for a single Big Five variable is larger than for two or five Big Five variables. Moreover, for larger sample sizes the standard deviation decreases especially for u_{kj} bias estimates.

In conclusion, the GLMM can estimate u_{kj} with acceptable bias values, bias and RMSE values are largest for a_j estimates and in some situations a larger sample size than 50 leads to lower standard deviation of the estimates.

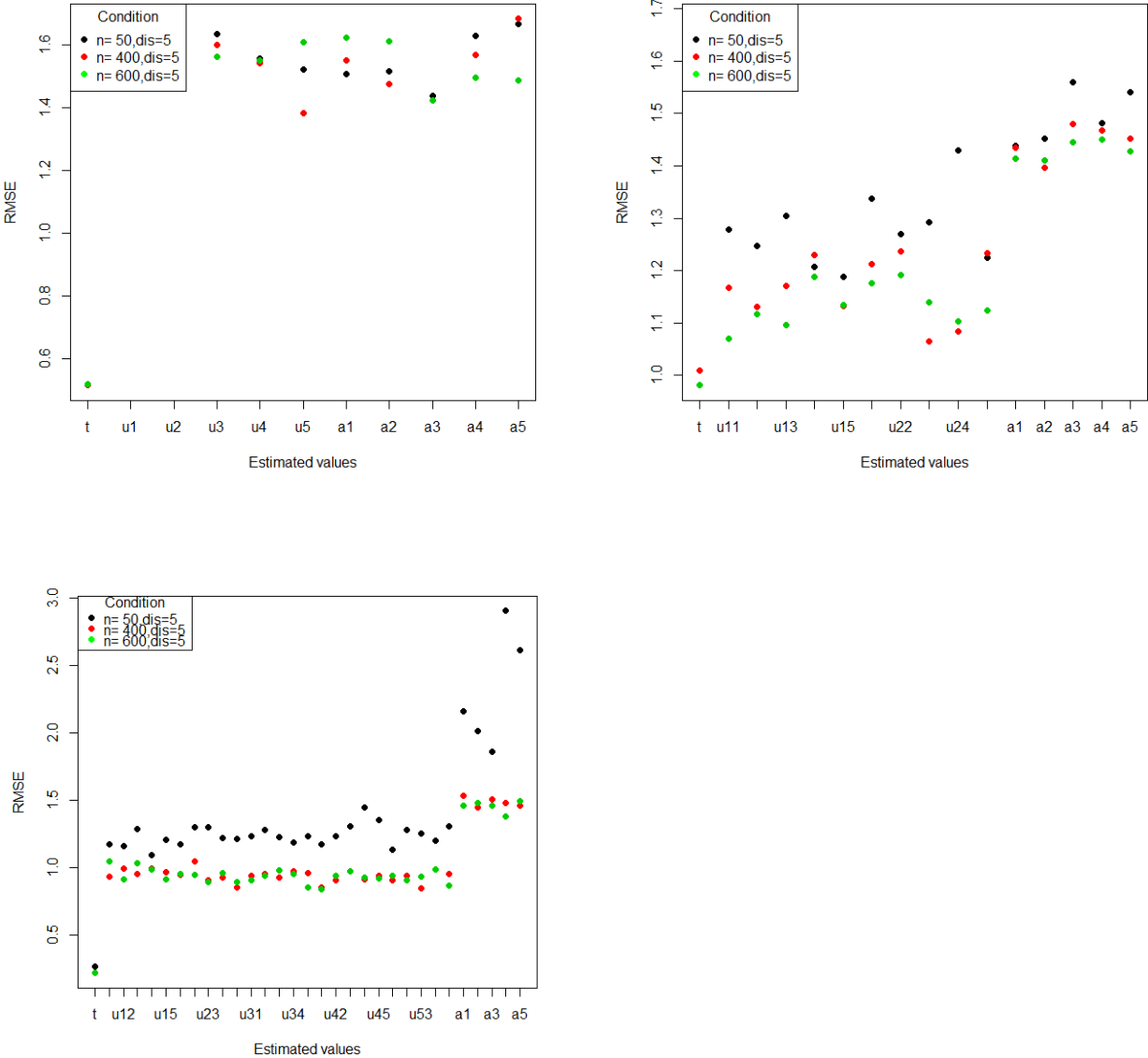


Figure 4 RMSE values for 1 Big Five variable (upper left), two Big Five variables (upper right) and five Big Five variables (bottom).

<i>bf</i>	<i>n</i>	<i>dis</i>	<i>t</i>	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{16}	u_{17}	u_{18}	u_{19}	u_{110}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	
1	50	5	0.51	0.01	0.15	0.26	0.03	0.05						1.18	1.03	1.20	1.35	1.18						
			10	0.50	0.00	0.17	0.06	0.01	0.22	0.03	0.10	0.10	0.10	0.16	1.06	1.13	1.23	1.20	1.31	1.27	1.40	1.03	1.29	1.04
	400	5	0.52	0.04	0.04	0.17	0.05	0.14						1.19	1.13	1.24	0.92	1.00						
			10	0.51	0.05	0.05	0.15	0.04	0.01	0.15	0.30	0.01	0.25	0.32	1.24	1.20	1.26	1.06	1.21	1.26	1.06	1.09	1.22	1.14
	600	5	0.52	0.11	0.01	0.06	0.17	0.47							1.16	1.23	0.92	1.02	1.18					
			10	0.52	0.01	0.09	0.04	0.07	0.11	0.10	0.33	0.04	0.03	0.02	1.04	1.13	1.35	1.32	1.05	1.16	1.14	1.15	1.12	1.39
2	50	5	<i>d</i>	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{16}	u_{17}	u_{18}	u_{19}	u_{110}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	
			1.31	0.02	0.05	0.02	0.04	0.04								1.31	1.30	1.25	1.22	1.20				
	400	5	0.96	0.01	0.20	0.00	0.04	0.08							1.09	1.30	1.11	1.08	1.25					
			10	0.93	0.01	0.04	0.13	0.24	0.12	0.14	0.09	0.09	0.12	0.05	1.27	1.27	1.21	1.25	1.28	1.24	1.29	1.23	1.23	1.24
	600	5	0.94	0.01	0.01	0.04	0.17	0.07							1.31	1.27	1.27	1.22	1.18					
			10	0.97	0.07	0.07	0.01	0.15	0.14	0.02	0.11	0.01	0.17	0.01	1.26	1.22	1.29	1.29	1.16	1.28	1.14	1.29	1.29	1.29
5	50	5	<i>d</i>	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{16}	u_{17}	u_{18}	u_{19}	u_{110}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	
			0.26	0.05	0.11	0.11	0.12	0.01								1.85	1.67	1.81	1.87	1.80				
	400	5	0.22	0.02	0.06	0.10	0.08	0.08							1.44	1.33	1.35	1.35	1.41					
			10	0.21	0.04	0.14	0.08	0.08	0.11	0.01	0.03	0.11	0.19	0.05	1.37	1.36	1.33	1.37	1.34	1.28	1.35	1.41	1.39	1.39
	600	5	0.22	0.13	0.01	0.05	0.03	0.05							1.39	1.41	1.37	1.37	1.43					
			10	0.21	0.08	0.07	0.14	0.00	0.07	0.06	0.09	0.04	0.07	0.06	1.40	1.29	1.36	1.39	1.33	1.35	1.33	1.38	1.28	1.40

Table 3 Bias values (red is negative): *bf* = number of Big Five variables, *n* = sample size, *dis* = number of disorders, *t* = tolerance, u_{1j} = Big Five variable value for top of disorder *j*, a_j = maximum probability of occurrence disorder *j*

<i>bf</i>	<i>n</i>	<i>dis</i>	<i>t</i>	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{16}	u_{17}	u_{18}	u_{19}	u_{110}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	
1	50	5	0.52	1.62	1.67	1.49	1.58	1.62						1.56	1.62	1.65	1.64	1.55						
		10	0.51	1.53	1.50	1.61	1.59	1.60	1.58	1.56	1.64	1.57	1.58	1.56	1.53	1.48	1.54	1.56	1.47	1.62	1.52	1.59	1.55	
	400	5	0.52	1.59	1.55	1.55	1.67	1.52							1.57	1.53	1.54	1.40	1.69					
		10	0.52	1.58	1.27	1.42	1.58	1.48	1.47	1.55	1.58	1.38	1.43	1.34	1.65	1.47	1.39	1.52	1.51	1.49	1.38	1.54	1.48	
	600	5	0.52	1.59	1.53	1.56	1.50	1.61							1.46	1.59	1.64	1.60	1.52					
		10	0.52	1.49	1.62	1.46	1.46	1.40	1.53	1.55	1.42	1.54	1.52	1.56	1.45	1.53	1.57	1.58	1.46	1.49	1.61	1.46	1.53	
				<i>d</i>	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{16}	u_{17}	u_{18}	u_{19}	u_{110}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
2	50	5	1.43	1.37	1.24	1.23	1.31	1.36						1.32	1.48	1.56	1.38	1.41						
		10	2.06	1.16	1.13	1.20	1.34	1.30	1.35	1.23	1.14	1.23	1.28	1.47	1.40	1.53	1.31	1.39	1.39	1.46	1.51	1.44	1.41	
	400	5	0.95	1.09	1.09	1.13	1.15	1.19							1.42	1.32	1.40	1.35	1.36					
		10	1.00	1.14	1.04	1.14	1.12	1.07	1.11	1.00	1.17	1.10	1.17	1.35	1.32	1.39	1.41	1.34	1.36	1.43	1.29	1.42	1.41	
	600	5	0.98	1.17	1.14	1.10	1.10	1.19							1.30	1.35	1.41	1.35	1.31					
		10	1.00	1.08	1.05	1.14	1.10	1.07	1.03	1.05	1.16	1.19	1.08	1.35	1.45	1.40	1.39	1.38	1.37	1.42	1.39	1.39	1.37	
				<i>D</i>	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{16}	u_{17}	u_{18}	u_{19}	u_{110}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
5	50	5	0.27	1.30	1.29	1.16	1.20	1.23						2.06	2.06	1.75	1.98	1.90						
		10	0.31	1.17	1.17	1.29	1.16	1.31	1.15	1.28	1.26	1.21	1.20	1.63	1.64	1.60	1.75	1.68	1.75	1.71	1.80	1.60	1.69	
	400	5	0.22	0.97	0.99	0.98	1.02	0.86							1.57	1.57	1.59	1.64	1.62					
		10	0.22	1.04	0.96	0.98	0.96	1.01	0.94	0.93	0.96	0.97	0.97	1.42	1.46	1.45	1.45	1.45	1.50	1.55	1.45	1.51	1.48	
	600	5	0.21	0.90	0.88	0.93	0.86	0.94							1.55	1.57	1.49	1.56	1.50					
		10	0.22	1.00	0.84	0.87	0.98	0.96	0.90	0.90	1.00	0.96	0.86	1.49	1.60	1.52	1.52	1.56	1.50	1.63	1.50	1.53	1.54	

Table 4 RMSE values: *bf* = number of Big Five variables, *n* = sample size, *dis* = number of disorders, *t* = tolerance, u_{1j} = Big Five variable value for top of disorder *j*, a_j = maximum probability of occurrence disorder *j*

3.2 Results NESDA data

To examine whether there is a unimodal relationship between the occurrence of anxiety and depressive disorders and the Big Five personality traits, generalized linear mixed models are fitted to the NESDA data. First of all, the Big Five personality factors are examined one by one. Again the test for unimodality as in equation (4) is executed. Table 5 shows the results of the hypothesis tests. For all Big Five personality factors, except for openness, there is a significant result, indicating rejection of the null hypothesis. So, for extraversion, conscientiousness, agreeableness and neuroticism there seems to be a unimodal relationship with the occurrence of anxiety and depressive disorders.

Big Five personality	<i>p</i> - value
Neuroticism	$7.8 \cdot 10^{-204} *$
Extraversion	$1.7 \cdot 10^{-92} *$
Openness	.3
Agreeableness	0.001*
Conscientiousness	$3.8 \cdot 10^{-21} *$

Table 5 *p*- values unimodality test (NESDA)

Figure 5 shows the relationship between the estimated person effects and the two Big Five variables neuroticism and openness. There are six different lines visible. The lines represent persons with different number of disorders. The bottom line represents 258 persons without a disorder. The upper line represents 5 persons with all five disorders. In between are persons with 1 ($n = 154$), 2 ($n = 96$), 3 ($n = 59$) and 4 ($n = 28$) disorders. Left the N-shaped relationship of neuroticism and the person effects graphically indicates a unimodal response. On the contrary openness (right) doesn't show an N-shaped relationship with the person effects. There is no indication for a unimodal response.

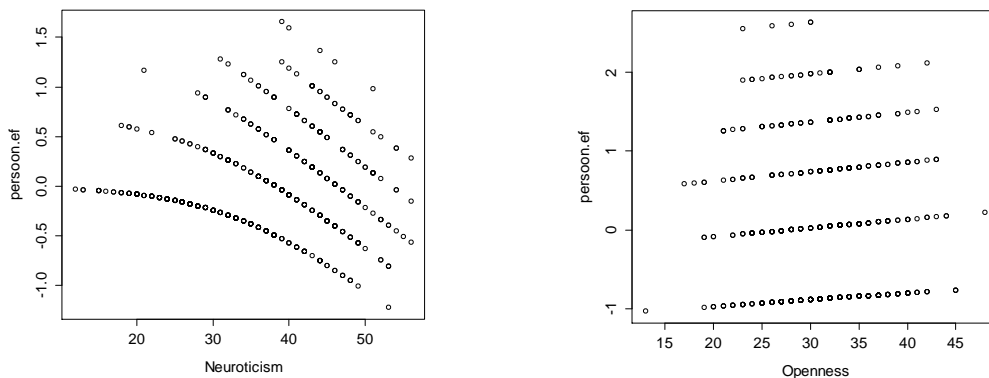


Figure 5 Estimated person effects GLMM against Big Five variable: neuroticism (left) and openness (right)

For each of the five disorders the values for u_j , a_j and t_j are computed separately for each Big Five variable, using equations (8). The results are presented in Table 6. The optimum cannot be estimated well if it lies outside of near the edge of the Big Five range, leading to $b_2 > 0$. That's why some t_j values are missing.

Big Five	Disorder	u_j	a_j	t_j
Neuroticism	Dysthymia	54.70	-12.95	.04
	Major depression	113.58	-7.56	.02
	Generalized anxiety	54.81	-13.57	.05
	Social phobia	53.21	-16.36	.05
	Panic disorder	50.07	-9.05	.04
Extraversion	Dysthymia	96.83	3.33	
	Major depression	67.15	7.53	
	Generalized anxiety	.81	-.07	.03
	Social phobia	19.62	-1.98	.05
	Panic disorder	10.14	-.07	.02
Openness	Dysthymia	28.00	-4.53	.04
	Major depression	42.86	.64	
	Generalized anxiety	29.66	-3.20	.03
	Social phobia	24.07	-2.21	.03
	Panic disorder	-1.58	-.68	.01
Agreeableness	Dysthymia	49.17	8.92	
	Major depression	64.60	4.53	
	Generalized anxiety	1.15	-.08	.02
	Social phobia	39.32	-11.21	.05
	Panic disorder	37.72	-4.71	.04
Conscientiousness	Dysthymia	51.85	5.94	
	Major depression	46.23	11.72	
	Generalized anxiety	22.11	-2.72	.04
	Social phobia	59.52	4.54	
	Panic disorder	79.08	2.03	

Table 6 NESDA data: u_j , a_j and t_j values

u_j represents the Big Five value for which the appearance of the disorder is most likely. Low u_j values mean that the disorders' optimum is reached for very low levels on the personality trait. For example generalized anxiety disorder often has a low u_j value. The disorders' optimum is reached for Big Five values outside the range. High u_j values mean that the disorders' optimum is reached for very high levels on the personality trait. All disorders have high u_j values for Big Five variable neuroticism. The disorders' top is reached for unrealistic high scores on neuroticism. a_j relates to the maximum probability of occurrence. The higher a_j , the higher the maximum probability of occurrence of the disorder. Often the maximum probability of occurrence of major depression is highest. t_j corresponds to the width of the curve. For all disorders the tolerance is roughly equal.

Figure 6 shows the unimodal curves for the five Big Five variables. As can be seen, the maximum probabilities of occurrence with respect to neuroticism are reached at the higher end of the scale. None of the NESDA participants had such a high score. For extraversion the top of the curves are at the lower end of the extraversion continuum. For openness there isn't a unimodal relationship. For agreeableness and conscientiousness the maximum probabilities of occurrence lie more in the middle of the Big Five variable. For example, low conscientiousness scores lead to high probability of occurrence of major depression disorder. The higher the conscientiousness score, the lower the probability of occurrence of major depression disorder. For approximately a conscientiousness score of 45 the probability of occurrence of major depression disorder reaches his lowest point. For higher scores on conscientiousness, the probability increases. In this case, the unimodal relationship is assumable.

Next, all Big Five personality factors are added to the generalized linear mixed model at once. In comparison with the simulation study, the variance of the predictors can differ ($d_1 \neq d_2 \neq d_3 \neq d_4 \neq d_5$) and it is unrealistic to assume that there are no covariances ($d_{12j}, d_{13j} \dots \neq 0$) between the predictors. But, we do assume the covariances to be equal for all disorders ($d_{12j} = d_{12}, d_{13j} = d_{13} \dots$). Jamil and Ter Braak (2013) denote that when this is the case, the model can do without interactions, because those can be subsumed in to the person effects (γ_i). The statistical test for unimodality includes both main, interaction and squared effects. This gives a significant result ($p < .001$). The alternative model with squared terms is accepted, indicating unimodal responses.

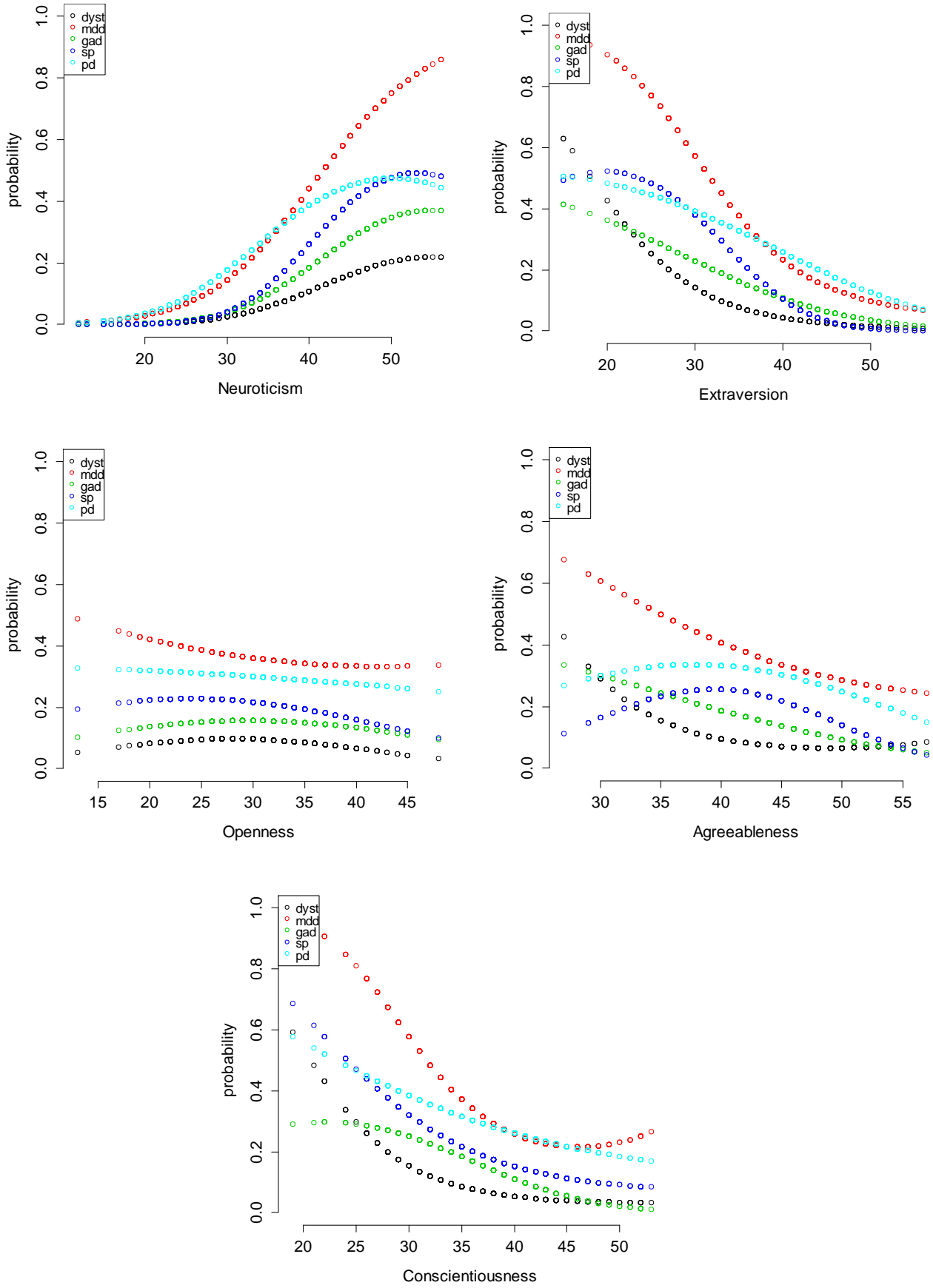


Figure 6 NESDA: Unimodal curves

4. Discussion

Often linear analysis methods are used to analyze psychological data. However, more and more interest for non-linear models emerges (Mattei, 2014). Jamil and Ter Braak (2013) used generalized linear mixed models to detect unimodal species- environmental relationships. This study generalizes the use of generalized linear mixed models to psychological context. It is illustrated with the NESDA data. The study is divided into two parts. The first part, the simulation study, evaluates the appropriateness of the generalized linear mixed model in identifying a unimodal relationship. In the second part the generalized linear mixed model is fitted to the NESDA data to examine whether there is a unimodal relationship between the occurrence of anxiety and depressive disorders and the Big Five personality traits.

4.1 The simulation study

In the simulation study, the ecological study of Jamil and ter Braak (2013) is the starting-point. As Jamil and ter Braak show, the unimodal Gaussian logistic curve can be written as a logistic linear mixed model. In this study the logistic linear mixed model is extended to five predictor variables. Moreover, the influence of sample size and number of disorders is examined. The generalized linear mixed model is evaluated by an analysis of power and an analysis of bias.

In the power analysis the statistical test for unimodality is executed multiple times to all experimental conditions. The power index shows the proportion of times the null hypothesis is rejected. The statistical test for unimodality shows very high power in all conditions. So the generalized linear mixed model is able to detect a unimodal relationship in data with the simulated characteristics.

In addition, a bias analysis examines bias values and the root-mean-square-errors. The bias values for the maximum probability of occurrence of a disorder (a_j) are relatively large. This indicates that the model has difficulties estimating the a_j 's or the "back" calculation for a_j isn't fully appropriate. For the calculation of a_j both the estimated u_{kj} and t/d are needed. This could lead to an incorrect estimate. Sample size does not have a considerable influence on the u_{kj} bias estimates. The GLMM is capable of estimating the parameters for a sample size of 50. The RMSE values, however, show a decline as the sample size increases. This indicates that the standard deviation of the estimates decreases as the sample size increases. Jamil and ter Braak (2013) simulated their unimodal data with one environmental variable, 10

or more species (corresponding to disorders) and 20-200 sites (corresponding to persons). The condition with a single Big Five variable, 10 disorders and sample size 50 approximates this situation. The other conditions, which are more like the NESDA data, show an equal or slightly higher bias and equal RMSE values. In the NESDA data set we use data from 600 persons. In reality almost 3000 persons participated to NESDA. Our results indicate that the generalized linear mixed model can deal with larger sample sizes. So the generalized linear mixed model seems to describe the data with NESDA characteristics equally well in comparison to the data of Jamil and Ter Braak (2013).

4.2 NESDA analysis

The second part of the present study is to examine whether there is a unimodal relationship between the occurrence of a disorder and the Big Five personality traits. Therefore the generalized linear mixed model is fitted to the NESDA data. Previous research reported strong positive correlations of for example neuroticism with symptoms of depression and anxiety and negative correlations of extraversion to depression (e.g. Jylhä & Isometsä, 2006). In this study we also find higher probabilities of occurrence of depression and anxiety disorders for higher levels of neuroticism and lower levels of extraversion.

Nowadays often linear analysis methods are used, but for all personality factors, except for openness, the test for unimodality gives a significant result. However, further inspection of the unimodal relationship between the occurrence of depression and anxiety disorders and the Big Five variables leads to more specific conclusions. For neuroticism and extraversion the maximum probability of occurrence lies at the extreme ends of the continuum. In practice, these neuroticism and extraversion scores do not occur. In this case the relation is more monotone than unimodal. The maximum probability of occurrences of disorders with respect to agreeableness and conscientiousness lie in the middle of the continuum. In that case, a unimodal relationship should be taken into serious consideration. This study gives indications that a non-linear model, which allows a unimodal relationship, is more appropriate than linear models in describing relations with two of the personality traits.

4.3 The fit of the GLMM and assumptions

Fitting the generalized linear mixed model can lead to several issues. In some of the replications, the person effects are estimated as 0, resulting in an immeasurable p -value for the statistical test. This occurs the most in the condition with the smallest sample size. In approximately 30% of the replications the random person effects are absent. In the condition

with larger sample sizes, up to 10% of the replications the p -value is immeasurable. In the condition with five Big Five variables the p -value is always measurable. The faulty replications were excluded from the analysis.

There were some struggles in the “back” calculation of the estimates. Sometimes the person effects were estimated as a positive value which made the calculation impossible. We had to take the median of the estimated t values in order to continue the calculation and with five personality factors we had to use a different precision parameter d . All these interventions might have minor influences on the outcome.

Moreover, for the data simulation a couple of assumptions are made. First of all, we assume equal tolerances for all disorders. If the tolerances would vary among disorders, the logistic linear mixed model does not hold because the random person effects (γ_i) would also depend on j . This is to be taken into consideration, because some disorders might be more general and have broader unimodal curves for specific Big Five variables. In this study the equi-tolerance assumption is not tested.

Secondly, with more than one Big Five variable we assume the variance of the predictors to be equal ($d_1 = d_2 = d_3 = d_4 = d_5$) and the absence of covariances between predictors ($d_{12j}, d_{13j} \dots = 0$). So the spread for all disorders is equal and there is no relation between the predictors. In practice these assumptions may not hold.

In conclusion, the results of the simulation study point out that a generalized linear mixed model is able to detect a unimodal relationship with high power. The bias values do not much differ from bias values from the condition with characteristics of the ecological data of Jamil and Ter Braak (2013).

4.4 Implications

Future research in personality assessment should more often focus on non-linear models. This study shows the usefulness of generalized linear mixed models in analyzing unimodal data. However, the GLMM showed some issues with convergence. The use of a different optimizer for the model may help. For the analysis of the NESDA data an analysis with the optimizer “bobyqa” instead of the default “Nelder_Mead” improves convergence. However, the differences between the two optimizers are negligible. We make assumptions about the variances and covariances of the predictors. More research is needed on the fit of generalized linear mixed models in the context of psychological data or, more general on models which allow a unimodal relationship. Results of linear models can be compared with the results of

unimodal models to obtain more insight into the structure of personality data. In the power analysis we focused on type II errors (concluding there is no effect, when, in fact there is), but in the hypothesis test a type I error (concluding there is an effect, when, in fact there is not) can occur as well. Future research can focus on this question.

Nowadays the majority of personality scales is developed with Likert's approach. Likert's approach implicitly assumes a dominance (monotonically increasing) response process. Items which are constructed with an ideal point response process underlying the response pattern (allowing unimodality) can also be included. As an extension to the present study, which demonstrate the importance of unimodality in personality, ideal point approaches should be considered to scale construction in personality assessment (Stark et al., 2006).

Altogether, the generalized linear mixed model is a promising method in identifying a unimodal relationship in data with psychological characteristics and more research is needed on unimodality in personality.

Bibliography

- Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1-19.
- Bienvenue, O.J., Samuels, J.F., Costa, P.T., Reti, M., Eaton, W.W., & Nestadt, G. (2004). Anxiety and depressive disorders and the five-factor model of personality: a higher- and lower- order personality trait investigation in a community sample. *Depression and Anxiety*, 20, 92-97.
- Cuijpers P., van Straten, A., & Donker, M. (2004). Personality traits of patients with mood and anxiety disorders. *Psychiatry Research*, 133, 229-237.
- Hayward, R.D., Taylor, W.D., Smoski, M.J., Steffens, D.C., & Payne, M.E. (2013). Association of five-factor model personality domains and facets with presence, onset, and treatment outcomes of major depression in older adults. *Journal of Geriatric Psychiatry*, 21, 88-96.
- Jamil, T., & ter Braak, C.J.F. (2013). Generalized linear mixed models can detect unimodal species-environment relationships. *PeerJ*, DOI 10.7717/peerj.95.
- Jamil, T., Kruk, C., & ter Braak, C.J.F. (2014). A unimodal species response model relating traits to environment with application to phytoplankton communities. *PloS One*, 9, 1-14.
- Jylhä, P., & Isometsä, E. (2006). The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depression and Anxiety*, 23, 281-189.
- Mattei, T.A. (2014). Unveiling complexity: non-linear and fractal analysis in neuroscience and cognitive psychology. *Frontiers in Computational Neuroscience*, 8, 1-2.
- Paxton, P., Curran, P.J., Bollen, K.A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: design and implementation. *Structural Equation Modeling*, 8, 287-312.
- Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., de Jong, P.J., van Marwijk, H.W.J., Assendelft, W.J.J., van der Meer, K., Verhaak, P., Wensing, M., de Graaf, R., Hoogendijk, W.J., Ormel, J. & van Dijck, R. (2008). The Netherlands study of depression and anxiety (NESDA): rationale, objectives and methods. *International Journal of Methods in Psychiatric Research*, 17, 121-140.
- Polak, M. G. (2011). *Item analysis of single peaked response data*. Rotterdam: Optima.
- Roberts, W.L. (1986). Nonlinear models of development: An example from the socialization of competence. *Child Development*, 57, 1166-1178.

- Roberts, B.W., Kuncel, N.R., Shiner, R., Caspi, A., & Goldberg, L.R. (2007). The power of personality. The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on psychological science*, 2, 313-345.
- Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35, 137-167.
- Spinhoven, P., Elzinga, B.M., van Hemert, A.M., de Rooij, M., & Penninx, B.W.J.H. (2014). A longitudinal study of facets of extraversion in depression and social anxiety. *Personality and Individual Differences*, 71, 39-44.
- Spinhoven, P., de Rooij, M., Heiser, W., Smit, J.H., & Penninx, B.W.J.H. (2009). The role of personality in comorbidity among anxiety and depressive disorders in primary care and specialty care: a cross-sectional analysis. *General Hospital Psychiatry*, 31, 470-477.
- Stark, S., Chernyshenko, O.S., Drasgow, F., & Williams, B.A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91, 25-39.

Appendix 1

A logistic linear mixed model can be seen as a Gaussian logistic model with equal tolerances. The unimodal model with more than one personality variable can be rewritten as a linear mixed model. No higher order interaction effects appear; only pairwise interaction effects. This is illustrated with three personality variables. d 's are precision parameters, corresponding to $1/t_j$. The Gaussian logistic model for three variables is as follows

$$\text{logit}(p_{ij}) = a_j - \frac{1}{2}(d_1(x_{i1} - u_{1j})^2 + d_2(x_{i2} - u_{2j})^2 + d_3(x_{i3} - u_{3j})^2 - 2d_{12j}(x_{i1} - u_{1j})(x_{i2} - u_{2j}) - 2d_{13j}(x_{i1} - u_{1j})(x_{i3} - u_{3j}) - 2d_{23j}(x_{i2} - u_{2j})(x_{i3} - u_{3j}))$$

By expanding the quadratic terms we can write

$$\alpha_j = a_j - \frac{1}{2}(d_1u_{1j}^2 + d_2u_{2j}^2 + d_3u_{3j}^2 - 2d_{12j}u_{1j}u_{2j} - 2d_{13j}u_{1j}u_{3j} - 2d_{23j}u_{2j}u_{3j})$$

$$\beta_{1j} = d_1u_{1j} - d_{12j}u_{2j} - d_{13j}u_{3j}$$

$$\beta_{2j} = d_2u_{2j} - d_{12j}u_{1j} - d_{23j}u_{3j}$$

$$\beta_{3j} = d_3u_{3j} - d_{13j}u_{1j} - d_{23j}u_{2j}$$

$$\beta_{4j} = d_{12j}$$

$$\beta_{5j} = d_{13j}$$

$$\beta_{6j} = d_{23j}$$

$$\gamma_i = -\frac{1}{2}(d_1x_{i1}^2 + d_2x_{i2}^2 + d_3x_{i3}^2)$$

and

$$\text{logit}(p_{ij}) = \alpha_j + \beta_{1j}x_{i1} + \beta_{2j}x_{i2} + \beta_{3j}x_{i3} + \beta_{4j}x_{i1}x_{i2} + \beta_{5j}x_{i1}x_{i3} + \beta_{6j}x_{i2}x_{i3} + \gamma_i$$

Extension to five personality variables is immediate. The Gaussian logistic model for all Big Five personality variables is the following

$$\begin{aligned} \text{logit}(p_{ij}) = & a_j - \frac{1}{2}(d_1(x_{i1} - u_{1j})^2 + d_2(x_{i2} - u_{2j})^2 + d_3(x_{i3} - u_{3j})^2 + d_4(x_{i4} - u_{4j})^2 + \\ & d_5(x_{i5} - u_{5j})^2 - 2d_{12j}(x_{i1} - u_{1j})(x_{i2} - u_{2j}) - 2d_{13j}(x_{i1} - u_{1j})(x_{i3} - u_{3j}) - \\ & 2d_{14j}(x_{i1} - u_{1j})(x_{i4} - u_{4j}) - 2d_{15j}(x_{i1} - u_{1j})(x_{i5} - u_{5j}) - 2d_{23j}(x_{i2} - u_{2j})(x_{i3} - u_{3j}) - \\ & 2d_{24j}(x_{i2} - u_{2j})(x_{i4} - u_{4j}) - 2d_{25j}(x_{i2} - u_{2j})(x_{i5} - u_{5j}) - 2d_{34j}(x_{i3} - u_{3j})(x_{i4} - u_{4j}) - \\ & 2d_{35j}(x_{i3} - u_{3j})(x_{i5} - u_{5j}) - 2d_{45j}(x_{i4} - u_{4j})(x_{i5} - u_{5j})) \end{aligned}$$

$$u_{3j}) - 2d_{24j}(x_{i2} - u_{2j})(x_{i4} - u_{4j}) - 2d_{25j}(x_{i2} - u_{2j})(x_{i5} - u_{5j}) - 2d_{34j}(x_{i3} - u_{3j})(x_{i4} - u_{4j}) - 2d_{35j}(x_{i3} - u_{3j})(x_{i5} - u_{5j}) - 2d_{45j}(x_{i4} - u_{4j})(x_{i5} - u_{5j}))$$

We can write

$$\text{logit}(p_{ij}) = \alpha_j + \beta_{1j}x_{i1} + \beta_{2j}x_{i2} + \beta_{3j}x_{i3} + \beta_{4j}x_{i4} + \beta_{5j}x_{i5} + \beta_{6j}x_{i1}x_{i2} + \beta_{7j}x_{i1}x_{i3} + \beta_{8j}x_{i1}x_{i4} + \beta_{9j}x_{i1}x_{i5} + \beta_{10j}x_{i2}x_{i3} + \beta_{11j}x_{i2}x_{i4} + \beta_{12j}x_{i2}x_{i5} + \beta_{13j}x_{i3}x_{i4} + \beta_{14j}x_{i3}x_{i5} + \beta_{15j}x_{i4}x_{i5} + \gamma_i$$

and, assuming the absence of covariances between the predictors, we can write

$$\alpha_j = \alpha_j - \frac{1}{2}(d_1u_{1j}^2 + d_2u_{2j}^2 + d_3u_{3j}^2 + d_4u_{4j}^2 + d_5u_{5j}^2)$$

$$\beta_{1j} = d_1u_{1j}$$

$$\beta_{2j} = d_2u_{2j}$$

$$\beta_{3j} = d_3u_{3j}$$

$$\beta_{4j} = d_4u_{4j}$$

$$\beta_{5j} = d_5u_{5j}$$

$$\gamma_i = -\frac{1}{2}(d_1x_{i1}^2 + d_2x_{i2}^2 + d_3x_{i3}^2 + d_3x_{i4}^2 + d_3x_{i5}^2)$$

In the lme4 package the model with five personality disorders can be fitted by

$$\text{lmer}(y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5 + (1 + x_1 + x_2 + x_3 + x_4 + x_5 | \text{disorder}) + (1 | \text{person}))$$