The robustness of the bivariate ordinal probit model to the presence of response styles

Rick Rijkse (343735)

Supervisor: Pieter Schoonees

July 1, 2013

Bachelor thesis
Econometrics & Operations Research
Erasmus University Rotterdam

Abstract

Respondents may give different answers to survey questions although they have the same underlying opinion. When data is collected among respondents with different response styles, interpretations of models based on this data could be wrong. In this project we describe shortly the idea of response styles. We will study, in the case of two ordinal variables, the extent to which the bivariate ordinal probit model is affected by the presence of response styles. First we explain the model and derive the log-likelihood function, we then show how this log-likelihood function can be maximized. We will apply this model to empirical data. Furthermore, we will simulate ordinal data and scale a part of this data with response style curves. The parameter estimates of the models based on the partly scaled and non-scaled data are compared to measure the robustness of this model to the presence of response styles. When (dis)acquiescence is present, the correlation coefficient is overestimated, the absolute value of the β coefficients decreases and the threshold values are shifted. This problem can be solved by adding a dummy-variable that indicates which respondents have this response style. The presence of midpoint responding and extreme responding leads to a change in the dispersion of the threshold values.

Contents

1 Introduction to response styles **4**

2 Bivariate ordinal probit model **5**
- 2.1 Motivation 5
- 2.2 Notation 5
- 2.3 Maximum likelihood estimation 6
- 2.4 Newton-Raphson algorithm 7
- 2.5 First-order derivatives of the log-likelihood function 8
- 2.6 Second-order derivatives of the log-likelihood function 10

3 Implementation **11**

4 Application **12**
- 4.1 Watching TV coverage of the war in Iraq 12
- 4.2 Parameter estimates and model interpretation 12

5 Robustness to response styles **14**
- 5.1 Simulation of response styles 14
- 5.2 Effects of response styles on parameter estimates 15

6 Conclusion **18**

7 Discussion **19**

References **20**

Appendix **21**
- A Recovering univariate ordered probit parameter estimates 21
- B Smooth density plots of simulation parameter estimates 22
- C Applying (dis)acquiescence to a known group 29

1 Introduction to response styles

There are different types of questions that are asked in surveys. One possible type is asking the respondents to map their opinions onto a Likert scale. For example, there are five ordered answer options, in a range from *strongly disagree* to *strongly agree*. We assume that the respondents have an underlying latent preference and subsequently transform their preference to one of the possible answer options. It is possible that respondents that have the same preference choose for different answer options. In other words, respondents could have different response styles. Some respondents would avoid the lowest and highest answer options; however other respondents tend toward often using the extreme answer options. Van de Velden (2007) defines a response style as a possibly non-linear mapping of the underlying latent preferences to a rating scale, that is common among a group of individuals. For example, response styles may differ across cultures, nationalities, education level, age, etc. The underlying latent preferences are unobservable, therefore it is difficult to detect a response style. Schoonees, van de Velden & Groenen (2013) have distinquished four different main response styles, which are visualized in Figure 1:

- *Acquiescence:* the upper part of the rating scale is used often
- *Disacquiescence:* the lower part of the rating scale is used often
- *Midpoint responding:* the middle categories are chosen often
- *Extreme responding:* the endpoints of the rating scale are chosen often

Response styles can induce spurious correlations. Most often the data is analyzed without taking response styles into account. Therefore spurious correlations will be mistakenly interpreted as being meaningful. One model for ordinal multivariate data which accounts for correlations is the multivariate ordered probit model. In this project we will study, in the case of two ordinal variables, the extent to which the bivariate ordinal probit model is affected by the presence of response styles. Firstly, we implement the bivariate ordinal probit model in R [10]. Thereafter we investigate the robustness of this model to the presence of different response styles through a simulation study.

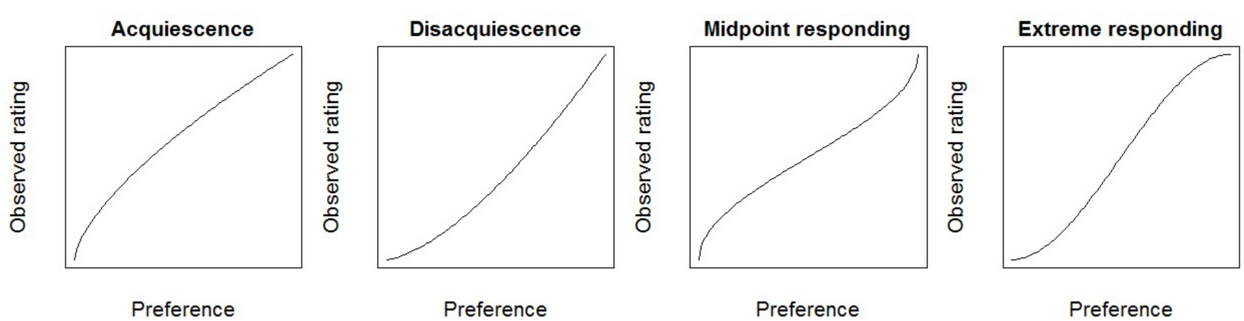

Figure 1: Different types of response styles

2 Bivariate ordinal probit model

In this section we explain the bivariate ordered probit model and derive the log-likelihood function and its first- and second-order derivatives. Furthermore, we explain how we could use the Newton-Raphson algorithm to estimate this model.

2.1 Motivation

To describe the data obtained by Likert-scale questions in surveys, we prefer a model that takes into account the ordering of the answer options. Therefore we don't use a multinomial probit model. When there are multiple dependent variables, often multiple univariate ordered probit models are used. However, there might be correlation between those variables. When we take the correlation into account, we expect a better model performance. The bivariate ordinal probit model meets this requirements for two dependent variables. Ashford & Sowden (1970) and Lesaffre & Molenberghs (1991) describe the multivariate probit model, however they do not pay much attention to ordered variables. In this section, the bivariate ordered probit model is explained and the log-likelihood function of this model is derived.

2.2 Notation

We define the ordinal variables Y_{i1} and Y_{i2} for $j = 1, 2, \ldots, J$ and $l = 1, 2, \ldots, L$ as

$$\begin{array}{lll} Y_{i1} = j & \text{iff} & \alpha_{j-1,1} < Y_{i1}^* \leq \alpha_{j,1} \\ Y_{i2} = l & \text{iff} & \alpha_{l-1,2} < Y_{i2}^* \leq \alpha_{l,2} \end{array} \tag{1}$$

where $\alpha_{0,1} < \alpha_{1,1} < \cdots < \alpha_{J,1}$ and $\alpha_{0,2} < \alpha_{1,2} < \cdots < \alpha_{L,2}$ are unobserved threshold values and Y_{ik}^* is a latent variable that represents the preference of respondent i for item k. Note that J is not necessarily equal to L. There are no upper and lower bounds to the latent variable, therefore we set $\alpha_{0,1} = \alpha_{0,2} = -\infty$ and $\alpha_{J,1} = \alpha_{L,2} = +\infty$.

We describe Y_{ik}^* with a bivariate regression model without intercepts

$$\left\{ \begin{array}{lll} Y_{i1}^* & = & \beta_1 x_i + \varepsilon_{i1} \\ Y_{i2}^* & = & \beta_2 x_i + \varepsilon_{i2} \end{array} \right. \tag{2}$$

in the case of only one explanatory variable x_i, where Y_{i1}^* and Y_{i2}^* are latent variables and ε_{i1} and ε_{i2} are continuous random variables. Note that there might be correlation between ε_{i1} and ε_{i2}. We impose without loss of generality that $\mathrm{Var}(\varepsilon_1) = \mathrm{Var}(\varepsilon_2) = 1$. We assume that ε_{i1} and ε_{i2} are bivariate normally distributed with mean zero and correlation ρ.

When there are more explanatory variables, we can write this model in matrix notation as

$$Y^* = X\beta + \varepsilon \tag{3}$$

where Y^* and ε are $N \times 2$ matrices, X is a $N \times V$ matrix and β is a $V \times 2$ matrix, where V is the number of explanatory variables.

2.3 Maximum likelihood estimation

The likelihood function is given by

$$L(\theta) = \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l=1}^{L} P[Y_{i1} = j, Y_{i2} = l|X_i]^{I(Y_{i1}=j, Y_{i2}=l)} \tag{4}$$

where θ is a summary of the parameters $\boldsymbol{\alpha}_1 = (\alpha_{0,1}, \alpha_{1,1}, \ldots, \alpha_{J,1})'$,
$\boldsymbol{\alpha}_2 = (\alpha_{0,2}, \alpha_{1,2}, \ldots, \alpha_{L,2})'$, $\boldsymbol{\beta}_1 = (\beta_{1,1}, \beta_{2,1}, \ldots, \beta_{V,1})'$, $\boldsymbol{\beta}_2 = (\beta_{1,2}, \beta_{2,2}, \ldots, \beta_{V,2})'$ and ρ.

The log-likelihood function is given by

$$l(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l=1}^{L} I(Y_{i1} = j, Y_{i2} = l) \log(P[Y_{i1} = j, Y_{i2} = l|X_i]) \tag{5}$$

where we can use that

$$\begin{aligned}
& P[Y_{i1} = j, Y_{i2} = l|X_i] \\
= \;& P[\alpha_{j-1,1} < Y_{i1}^* \leq \alpha_{j,1}, \alpha_{l-1,2} < Y_{i2}^* \leq \alpha_{l,2}] \\
= \;& P[\alpha_{j-1,1} < \mathbf{x}_i'\boldsymbol{\beta}_1 + \varepsilon_{i1} \leq \alpha_{j,1}, \alpha_{l-1,2} < \mathbf{x}_i'\boldsymbol{\beta}_2 + \varepsilon_{i2} \leq \alpha_{l,2}] \\
= \;& P[\alpha_{j-1,1} - \mathbf{x}_i'\boldsymbol{\beta}_1 < \varepsilon_{i1} \leq \alpha_{j,1} - \mathbf{x}_i'\boldsymbol{\beta}_1, \alpha_{l-1,2} - \mathbf{x}_i'\boldsymbol{\beta}_2 < \varepsilon_{i2} \leq \alpha_{l,2} - \mathbf{x}_i'\boldsymbol{\beta}_2] \\
= \;& \int_{\alpha_{j-1,1}-\mathbf{x}_i'\boldsymbol{\beta}_1}^{\alpha_{j,1}-\mathbf{x}_i'\boldsymbol{\beta}_1} \int_{\alpha_{l-1,2}-\mathbf{x}_i'\boldsymbol{\beta}_2}^{\alpha_{l,2}-\mathbf{x}_i'\boldsymbol{\beta}_2} \phi_2(\varepsilon_{i1}, \varepsilon_{i2}, \rho)\, \mathrm{d}\varepsilon_{i1}\mathrm{d}\varepsilon_{i2}
\end{aligned} \tag{6}$$

Here ϕ_2 is the bivariate normal density function with mean 0, variance 1 and correlation ρ.

2.4 Newton-Raphson algorithm

We can maximize the log-likelihood function using an iterative algorithm and implement this using the open-source statistical software R. According to Schwrorer & Hovey (2004), algorithms that can be used to maximize log-likelihood functions numerically are the Newton-Raphson method and the Fisher Scoring algorithm. In this project we will use the Newton-Raphson algorithm.

The maximum of the log-likelihood function can be found by applying

$$\theta_{h+1} = \theta_h - H(\theta_h)^{-1} G(\theta_h) \tag{7}$$

until convergence is reached, where $G(\theta_h)$ is the gradient evaluated in θ_h and $H(\theta_h)$ is the Hessian matrix evaluated in θ_h. This means that starting values θ_0 are needed. In the following subsections the first- and second-order derivatives of the log-likelihood with respect to θ are derived.

According to Efron & Hinkley (1978), to compute the standard errors of the parameter estimates we could use the observed information matrix, i.e. minus the second-order derivative of the log-likelihood function. The parameter estimates are asymptotic normally distributed:

$$\hat{\theta}_{ML} \sim \mathcal{N}(\theta_0, H^{-1}) \tag{8}$$

To have a warm start for finding the global maximum of the log-likelihood function, we first estimate two univariate ordered probit models using the *polr* function in the R-package *MASS*, written by Venables & Ripley (2002). We use the parameter estimates of these models as starting values for the Newton-Raphson algorithm. Using these starting values, convergence is reached quite fast. The algorithm is said to have converged when the log-likelihood changes by a less than a small constant $\epsilon > 0$. The convergence of the model described in section 4 is shown in Figure 2. In 18 iterations there is an increase in the log-likelihood less than 10^{-5}.

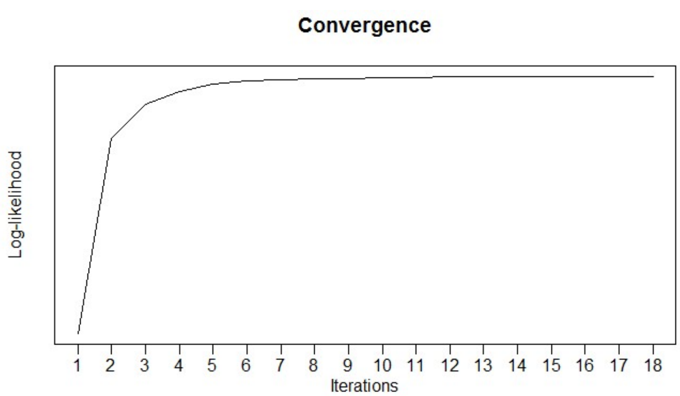

Figure 2: Example of convergence using the Newton-Raphson algorithm

We compute the multivariate integrals using the R-package *mvtnorm*. [7] The computational method of this package is described by Genz & Bretz (2009).

2.5 First-order derivatives of the log-likelihood function

The first-order derivatives of the log-likelihood function are given by

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L} \frac{I(Y_{i1}=j, Y_{i2}=l)}{P[Y_{i1}=j, Y_{i2}=l|X_i]} \frac{\partial P[Y_{i1}=j, Y_{i2}=l|X_i]}{\partial \boldsymbol{\theta}} \tag{9}$$

To obtain these derivatives, we describe (6) in terms of the cumulative distribution function of ε, where $\varepsilon \sim \mathcal{N}(0, \Sigma)$. We denote this cumulative distribution function by F. For convenience, we write the last expression in (6) as $P[L_j < \varepsilon_{i1} \leq U_j, L_l < \varepsilon_{i2} \leq U_l]$. We compute this probability by first computing $F(U_j, U_l)$. This is equal to the probability that ε_{i1} is below U_j and ε_{i2} is below U_l.

Then we need to substract three disjoint probabilities:

1. ε_{i1} is below L_j and ε_{i2} is between L_l and U_l: $F(L_j, U_l) - F(L_j, L_l)$
2. ε_{i2} is below L_l and ε_{i1} is between L_j and U_j: $F(U_j, L_l) - F(L_j, L_l)$
3. ε_{i1} is below L_j and ε_{i2} is below L_l: $F(L_j, L_l)$

If we combine this, we can make the following substitution in (9)

$$P[Y_{i1}=j, Y_{i2}=l|X_i] = F(U_j, U_l) - F(L_j, U_l) - F(U_j, L_l) + F(L_j, L_l) \tag{10}$$

To compute partial derivatives of the log-likelihood function with respect to L_j, L_l, U_j and U_l, we can use Lemma 1.

Lemma 1
Let $F(A, B)$ be a bivariate cumulative distribution function where A and B are standard normal random variables with correlation ρ. Bertsekas & Tsitsiklis (2002) describe that the conditional probability density function of B given $A = a$ is the normal density function with mean ρA and variance $\sqrt{1-\rho^2}$

$$\frac{\partial F(A,B)}{\partial A} = \frac{\partial}{\partial A}\int_{-\infty}^{A}\int_{-\infty}^{B} \phi_2(s,t,\rho)\mathrm{d}s\mathrm{d}t = \int_{-\infty}^{B} \phi_2(A,t,\rho)\mathrm{d}t = \phi(A) F\left(\frac{B-\rho A}{\sqrt{1-\rho^2}}\right). \tag{11}$$

The derivative with respect to ρ is particularly simple:

Lemma 2
Plackett (1954) derived the formula for the partial derivative of the bivariate normal cumulative distribution function with respect to the correlation coefficient ρ

$$\frac{\partial F(A,B)}{\partial \rho} = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{A^2+B^2-2\rho AB}{2(1-\rho^2)}\right) = \phi_2(A,B,\rho). \tag{12}$$

Using (10), the chain rule and Lemma 1 we can easily derive the first-order derivatives of the log-likelihood function with respect to $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. To impose the monotone order of the threshold values, we estimate instead a transformation of these values

$$
\begin{aligned}
\delta_{1,k} &= \alpha_{1,k} \\
\delta_{j,k} &= \log(\alpha_{j,k} - \alpha_{j-1,k}) \quad \forall j > 1
\end{aligned} \tag{13}
$$

Instead of estimating ρ using Lemma 2, we estimate the additional parameter $r = \operatorname{arctanh}(\rho)$. We use this Fisher transformation to obtain ρ on the interval [-1,1]. The derivatives of the bivariate normal distribution function with respect to $\delta_{j,k}$ and r are given by

$$
\begin{aligned}
\frac{\partial F(A,B)}{\partial \delta_{j,k}} &= \frac{\partial F(A,B)}{\partial \alpha_{j,k}} \frac{\partial \alpha_{j,k}}{\partial \delta_{j,k}} \\
\frac{\partial F(A,B)}{\partial r} &= \frac{\partial F(A,B)}{\partial \rho} \frac{\partial \rho}{\partial r}
\end{aligned} \tag{14}
$$

where we use that

$$
\begin{aligned}
\frac{\partial(\alpha_{j,k} - \mathbf{x}_i' \boldsymbol{\beta}_k)}{\partial \alpha_{j,k}} &= 1 \\
\frac{\partial(\alpha_{j,k} - \mathbf{x}_i' \boldsymbol{\beta}_k)}{\partial \beta_k} &= -\mathbf{x}_i \\
\frac{\partial \alpha_{1,k}}{\partial \delta_{1,k}} &= 1 \\
\frac{\partial \alpha_{j,k}}{\partial \delta_{j,k}} &= \exp(\delta_{j,k}) \\
\frac{\partial \rho}{\partial r} &= \tfrac{4\exp(2r)}{(1+\exp(2r))^2}
\end{aligned} \tag{15}
$$

2.6 Second-order derivatives of the log-likelihood function

The second-order derivatives of the log-likelihood function are given by

$$\frac{\partial^2 l(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1 \boldsymbol{\theta}_2} = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l=1}^{L} \left(\frac{I(Y_{i1} = j, Y_{i2} = l)}{P[Y_{i1} = j, Y_{i2} = l|X_i]} \frac{\partial^2 P[Y_{i1} = j, Y_{i2} = l|X_i]}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_2} - \frac{1}{P[Y_{i1} = j, Y_{i2} = l|X_i]^2} \frac{\partial P[Y_{i1} = j, Y_{i2} = l|X_i]}{\partial \boldsymbol{\theta}_1} \frac{\partial P[Y_{i1} = j, Y_{i2} = l|X_i]}{\partial \boldsymbol{\theta}_2} \right). \tag{16}$$

We derive the partial second-order derivatives of the bivariate normal cumulative distribution function using Clairaut's theorem, the chain rule, equation 10, Lemma 1 and Lemma 2

$$\begin{aligned}
\frac{\partial^2 F(A,B)}{\partial A^2} &= -A\phi(A)F\left(\frac{B-\rho A}{\sqrt{1-\rho^2}}\right) - \rho\phi_2(A,B,\rho) \\
\frac{\partial^2 F(A,B)}{\partial A \partial \rho} &= \frac{\rho B - A}{1-\rho^2}\phi_2(A,B,\rho) \\
\frac{\partial^2 F(A,B)}{\partial \rho^2} &= \frac{AB + \rho + \frac{\rho(A^2+B^2-2\rho AB)}{1-\rho^2}}{1-\rho^2}\phi_2(A,B,\rho) \\
\frac{\partial^2 F(A,B)}{\partial A \partial B} &= \phi_2(A,B,\rho)
\end{aligned} \tag{17}$$

The second-order derivatives of the log-likelihood function with respect to $\delta^2_{j_k,k}$ and r^2 are given by

$$\begin{aligned}
\frac{\partial^2 F(A,B)}{\partial \delta^2_{j,k}} &= \frac{\partial^2 F(A,B)}{\partial \alpha^2_{j,k}} \left(\frac{\partial \alpha_{j,k}}{\partial \delta_{j,k}}\right)^2 + \frac{\partial F(A,B)}{\partial \alpha_{j,k}} \frac{\partial^2 \alpha_{j,k}}{\partial \delta^2_{j,k}} \\
\frac{\partial^2 F(A,B)}{\partial r^2} &= \frac{\partial^2 F(A,B)}{\partial \rho^2} \left(\frac{\partial \rho}{\partial r}\right)^2 + \frac{\partial F(A,B)}{\partial \rho} \frac{\partial^2 \rho}{\partial r^2}
\end{aligned} \tag{18}$$

where we use that

$$\begin{aligned}
\frac{\partial^2 \alpha_{j,k}}{\partial \delta^2_{j,k}} &= \exp(\delta_{j,k}) \\
\frac{\partial^2 \rho}{\partial r^2} &= \frac{8\exp(2r) - 8\exp(4r)}{(1+\exp(2r))^3}
\end{aligned} \tag{19}$$

3 Implementation

In this section we explain how we have implemented the bivariate ordered probit model in R and show some examples. We have created separate functions for computing the likelihood, the first-order derivatives and the second-order derivatives of the log-likelihood function. These are called by a function that gives an update of the parameters, which is called by a function that runs the Newton-Raphson algorithm. The implemented Newton-Raphson algorithm doesn't always provide a higher log-likelihood after each iteration before convergence is reached. Theoretically this should not happen and hence this behaviour is an artefact of the implementation (possibly due to numerical issues). It is also possible that the algorithm doesn't converge at all, then different starting values could be used to obtain the maximum of the log-likelihood function.

We have simulated data to see whether the model is able to recover the actual parameters we have used to simulate the data. We have used the following data generating process (DGP)

$$\begin{aligned} y_1 &= x_1 + x_2 + \varepsilon_1 \\ y_2 &= \tfrac{3}{2} x_2 + \varepsilon_2 \end{aligned} \tag{20}$$

where x_1 and x_2 are both standard uniformly distributed and ε_1 and ε_2 are joint standard normally distributed with correlation 0.5. In Table 1, we see that the parameter estimates are close to the actual parameter values.

parameter	actual value	estimate	standard deviation
ρ	0.5	0.52	0.08
$\beta_{1,1}$	1	1.08	0.11
$\beta_{2,1}$	1	1.05	0.10
$\beta_{1,2}$	0	0.02	0.08
$\beta_{2,2}$	1.5	1.50	0.12
$\alpha_{1,1}$	-1.4	-1.43	0.14
$\alpha_{2,1}$	-0.4	-0.45	0.10
$\alpha_{3,1}$	0.6	0.49	0.09
$\alpha_{4,1}$	1.5	1.32	0.12
$\alpha_{1,2}$	-1.4	-1.49	0.14
$\alpha_{2,2}$	-0.5	-0.41	0.10
$\alpha_{3,2}$	0.2	0.44	0.08
$\alpha_{4,2}$	1	0.97	0.09

Table 1: Recovering actual parameters ($n = 200$)

When we set the actual correlation to 0, we can compare the coefficients of the bivariate ordered probit model and two univariate ordered probit models. The parameter estimates should be the same and in Table 6 in Appendix A we see they almost are. When we increase the number of respondents, the parameter estimates are more close to one another.

4 Application

4.1 Watching TV coverage of the war in Iraq

In this section we will show a simple application of the bivariate ordered probit model. Suppose we are interested in how Americans were feeling when they viewed coverage of the war in Iraq on TV in 2003. We would like to know whether age and gender are influencing the sadness and fear of the people. We use data from the Pew Internet and American Life Project, which is collected by Princeton Survey Research Associates (2003).

The response variables are sadness (I feel sad when watching TV coverage of the war); and fear (It's frightening to watch TV coverage of the war). The possible answers to the response variables are strongly agree; agree; disagree; strongly disagree; and don't know/refused. The explanatory variables are age (18, 19, ..., 97, don't know, refused); and gender (0: male, 1: female).

We filter out the respondents that have answered *Don't know/Refused* to the response variables or to their age. There is data from 1390 respondents left. The response variables are ordered and it's likely that there is correlation between the two response variables, therefore we estimate the bivariate ordered probit model.

4.2 Parameter estimates and model interpretation

The parameter estimations of the latent variables are

$$\left\{\begin{array}{rcl} \text{Sadness}_i^* & = & -0.004\ (0.001)\ \text{Age}_i - 0.732\ (0.056)\ \text{Gender}_i + \varepsilon_{i1} \\ \text{Fear}_i^* & = & -0.003\ (0.001)\ \text{Age}_i - 0.707\ (0.053)\ \text{Gender}_i + \varepsilon_{i2} \\ \text{cor}(\varepsilon_{i1}, \varepsilon_{i2}) & = & 0.552\ (0.017) \end{array}\right. \tag{21}$$

where the standard deviations are in the brackets.

The estimated threshold values are

$$\left\{\begin{array}{lll} Sadness = \text{Strongly agree} & \text{iff} & Sadness_i^* \leq -1.546\ (0.047) \\ Sadness = \text{Agree} & \text{iff} & -1.546\ (0.047) < Sadness_i^* \leq -0.260\ (0.037) \\ Sadness = \text{Disagree} & \text{iff} & -0.260\ (0.037) < Sadness_i^* \leq 0.881\ (0.044) \\ Sadness = \text{Strongly disagree} & \text{iff} & Sadness_i^* > 0.881\ (0.044) \\ & & \\ Fear = \text{Strongly agree} & \text{iff} & Fear_i^* \leq -1.686\ (0.047) \\ Fear = \text{Agree} & \text{iff} & -1.686\ (0.047) < Fear_i^* \leq -0.492\ (0.037) \\ Fear = \text{Disagree} & \text{iff} & -0.492\ (0.037) < Fear_i^* \leq 0.804\ (0.043) \\ Fear = \text{Strongly disagree} & \text{iff} & Fear_i^* > 0.804\ (0.043) \end{array}\right. \tag{22}$$

Both coefficients for *Age* are negative, which means that older people feel more sad when they watch TV coverage of the war and are more frightened than younger people. Both coefficients for *Gender* are also negative. According to this model, men feel less sad and are less frightened than women when watching TV coverage of the war in Iraq. Note that this could be due to the response styles of men who might not want to admit that they are frightened or feel sad, because they think men shouldn't be.

Suppose we would like to predict the response of an American given his/her gender and age. The predicted response values are shown in Table 2.

Gender	Age	Sadness	Fear
Men	≤ 63	Disagree	Disagree
Men	> 63	Agree	Disagree
Women	all	Agree	Agree

Table 2: Predicted response values

The forecasting performance of this model can be found in Table 3. The hit rate for sadness is $(367 + 259)/1390 = 0.45$ and the hit rate for fear is $(292 + 358)/1390 = 0.47$.

Sadness	Predicted values	
Response values	Agree	Disagree
Strongly agree	182	55
Agree	**367**	231
Disagree	168	**259**
Strongly disagree	45	83

Fear	Predicted values	
Response values	Agree	Disagree
Strongly agree	129	45
Agree	**292**	212
Disagree	195	**358**
Strongly disagree	36	123

Table 3: Forecasting performance

Note that the predictive power of this model isn't very high. The model is not able to predict the possible answers *Strongly agree* and *Strongly disagree*. This also happens when we estimate univariate models based on this data. This does not mean this model is bad, since we only have used gender and age as explanatory variables. There could be a lot of more factors that have influence on sadness and fear of the respondents. When taking that variables into account, a better forecasting performance could be achieved.

5 Robustness to response styles

To measure the robustness of the bivariate ordinal probit model to response styles, we simulate ordinal data and scale a part (50%) of this data with the response style curves in Figure 3. We will estimate models based on the partly scaled data and compare these with models based on the non-scaled data as benchmark. We can compare the coefficients of the models. Furthermore, we can compare the mean squared errors and hitrates of both models.

5.1 Simulation of response styles

We have transformed the range of the simulated data $\hat{Y}$ to the range [0,1]. Then we have scaled this data with the following response style curves, which corresponds with the response style curves in Figure 3:

$$\begin{array}{ll} \text{No response style} & Y^* = \hat{Y} \\ \text{Acquiescence} & Y^* = \hat{Y}^{\frac{2}{3}} \\ \text{Disacquiescence} & Y^* = \hat{Y}^{\frac{3}{2}} \\ \text{Midpoint responding} & Y^* = I_{\hat{Y}}(0.5, 0.5) \\ \text{Extreme responding} & Y^* = I_{\hat{Y}}(2, 2) \end{array} \tag{23}$$

where $I_{\hat{Y}}(\theta_1, \theta_2)$ is the regularized incomplete beta function, which is given by

$$I_{\hat{Y}}(\theta_1, \theta_2) = \frac{\int_0^{\hat{Y}} t^{\theta_1 - 1}(1-t)^{\theta_2 - 1} dt}{\int_0^1 t^{\theta_1 - 1}(1-t)^{\theta_2 - 1} dt}. \tag{24}$$

Thereafter we have transformed Y^* back to the original range. For each response style, we have estimated a bivariate ordered probit model with 3 covariates (the first of those is a dummy-variable) and 5 possible response values corresponding with a Likert scale. We have simulated data for 200 respondents and have applied the response styles to 100 randomly selected respondents.

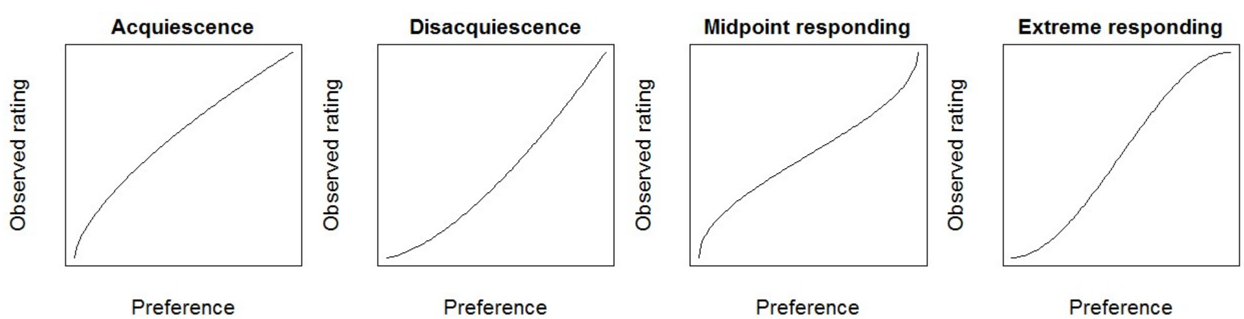

Figure 3: Response style curves

5.2 Effects of response styles on parameter estimates

First we have estimated the model where all 200 respondents don't have a response style. We compare the parameter estimates with two models where two opposite response styles are present: acquiescence (yea-saying) and disacquiescence (nay-saying). For each response style, the parameters of the model are estimated 100 times. The average parameter estimates and the estimated standard deviations of the parameter estimates can be found in Table 4.

parameter	actual value	no response style estimate	acquiescence estimate	disacquiescence estimate
ρ	0.5	0.51 (0.08)	0.73 (0.06)	0.73 (0.06)
$\beta_{1,1}$	1	1.04 (0.17)	0.75 (0.14)	0.77 (0.14)
$\beta_{2,1}$	1	1.02 (0.12)	0.75 (0.11)	0.78 (0.11)
$\beta_{3,1}$	-2	-2.05 (0.17)	-1.48 (0.15)	-1.52 (0.16)
$\beta_{1,2}$	0	0.03 (0.20)	-0.02 (0.14)	0.03 (0.13)
$\beta_{2,2}$	2	2.04 (0.17)	1.45 (0.14)	1.46 (0.16)
$\beta_{3,2}$	1.5	1.54 (0.13)	1.10 (0.12)	1.10 (0.12)
$\alpha_{1,1}$	-1.5	-1.54 (0.19)	-1.82 (0.16)	-0.38 (0.17)
$\alpha_{2,1}$	0	0.01 (0.15)	-0.72 (0.14)	0.71 (0.14)
$\alpha_{3,1}$	1	1.02 (0.15)	0.03 (0.16)	1.40 (0.16)
$\alpha_{4,1}$	2.5	2.58 (0.24)	1.24 (0.24)	2.48 (0.20)
$\alpha_{1,2}$	-2	-2.03 (0.22)	-2.21 (0.21)	-0.65 (0.15)
$\alpha_{2,2}$	-0.5	-0.48 (0.17)	-1.14 (0.16)	0.36 (0.14)
$\alpha_{3,2}$	0.5	0.53 (0.14)	0.38 (0.14)	1.03 (0.14)
$\alpha_{4,2}$	2	2.06 (0.19)	0.81 (0.19)	2.07 (0.19)

Table 4: (Dis)acquiescence simulation study ($n = 200$)

We see that:

- The correlation parameter estimate increases when a part of the population has a (dis)acquiescence response style.
- The β coefficients are multiplied with about $\frac{3}{4}$.
- The threshold value estimates are shifted down when acquiescence occurs, because the respondents tend to answer with higher ratings. When there is a disacquiescence response style present, the opposite holds: the threshold values are shifted upwards.

The parameter estimates of the midpoint responding and the extreme responding response styles, and the estimated standard deviations, can be found in Table 5.

parameter	actual value	no response style estimate	midpoint responding estimate	extreme responding estimate
ρ	0.5	0.51 (0.08)	0.44 (0.09)	0.47 (0.10)
$\beta_{1,1}$	1	1.04 (0.17)	0.97 (0.19)	1.00 (0.16)
$\beta_{2,1}$	1	1.02 (0.12)	0.97 (0.10)	0.99 (0.12)
$\beta_{3,1}$	-2	-2.05 (0.17)	-1.94 (0.19)	-1.98 (0.17)
$\beta_{1,2}$	0	0.03 (0.20)	-0.03 (0.19)	0.03 (0.20)
$\beta_{2,2}$	2	2.04 (0.17)	1.89 (0.16)	1.98 (0.17)
$\beta_{3,2}$	1.5	1.54 (0.13)	1.43 (0.14)	1.50 (0.15)
$\alpha_{1,1}$	-1.5	-1.54 (0.19)	-1.92 (0.25)	-1.17 (0.21)
$\alpha_{2,1}$	0	0.01 (0.15)	-0.14 (0.25)	0.10 (0.16)
$\alpha_{3,1}$	1	1.02 (0.15)	1.12 (0.31)	0.92 (0.16)
$\alpha_{4,1}$	2.5	2.58 (0.24)	2.88 (0.30)	2.18 (0.22)
$\alpha_{1,2}$	-2	-2.03 (0.22)	-2.37 (0.30)	-1.66 (0.23)
$\alpha_{2,2}$	-0.5	-0.48 (0.17)	-0.62 (0.27)	-0.42 (0.21)
$\alpha_{3,2}$	0.5	0.53 (0.14)	0.61 (0.26)	0.42 (0.18)
$\alpha_{4,2}$	2	2.06 (0.19)	2.36 (0.23)	1.66 (0.23)

Table 5: Midpoint responding and extreme responding simulation study ($n = 200$)

Here we conclude that:

- The correlation parameter estimate is close to the actual parameter value.
- The β coefficients are also estimated well.
- The threshold values estimates are more spread out when a part of the respondents tend to respond often with the midpart of the Likert scale. When there is a lot of extreme responding, the threshold values are less spread out.

We have computed smooth kernel density curves of the parameter estimates using the R package *sm*, written by Bowman & Azzalini (2013). In Figure 4, the density plot of the parameter estimates of ρ is shown. We can clearly see that when (dis)acquiescence is present, the correlation coefficient is estimated too high. The density plots of the other parameter estimates can be found in Appendix B.

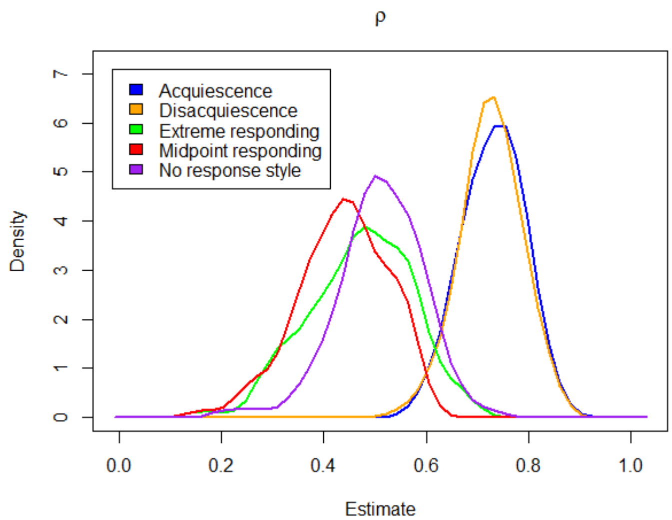

Figure 4: Density plot of parameter estimates of ρ (actual value = 0.5)

We also have applied the acquiescence and disacquiescence response styles to "a known part" of the respondents. The covariate x_1 is a dummy-variable that indicates whether a respondent belong to this group. The average parameter estimates and standard deviations can be found in Appendix C. Now we have different results:

- The correlation parameter estimate is estimated very well.
- The corresponding coefficients $\beta_{1,1}$ and $\beta_{1,2}$ are shifted downwards when acquiescence is present and shifted upwards when disacquiescence is present.
- The other β estimates are close to the actual parameter values.
- The threshold values are shifted down/up due to a change in the $\beta_{1,1}$ and $\beta_{1,2}$ estimates.

When we apply the midpoint responding and extreme responding response styles to a known part of the respondents, the results are the same as when we apply the response styles randomly. This is because there is a difference between the (dis)acquiescence response styles and the midpoint/extreme responding response styles: the first ones changes the skewness of the distribution of the simulated data, the last ones don't.

6 Conclusion

Surveys are often used by companies to measure consumer mindset metrics, for example brand associations, purchase intention and customer satisfaction. A problem with modelling survey data is that there could be bias due to the presence of response styles. When we analyze the bivariate ordered probit model while (dis)acquiescence is present, the correlation coefficient is estimated too high, the absolute value of the β coefficients decreases and the threshold values are shifted. This problem can be solved by adding a dummy-variable that indicates which respondents have this response style. Then the correlation coefficient and the β coefficients (except those who correspond with the dummy-variable) are estimated close to the actual values. The threshold values are still shifted. The presence of midpoint responding leads to an increase in the dispersion of the threshold values, extreme responding leads to a decrease.

7 Discussion

In this project, we have explored the robustness of the bivariate ordered probit model to the presence of response styles. These results are useful to take into account when response styles are present; however we didn't explore how one could detect response styles. That doesn't fall within the scope of this project, but is a really interesting topic for further research.

There are some limitations to the software implementation in this project:

- The Newton-Raphson algorithm doesn't always converge. When convergence is reached, the log-likelihood doesn't always increase monotonically. An example is shown in Figure 5.
- We use the parameter estimates of two univariate ordered probit models as "warm starts" for the Newton-Raphson algorithm. Using multiple random starting values may improve the chance of finding the global maximum of the log-likelihood function.
- In expression (10), we take differences of four integrals. These integrals are approximated. Instead of approximating four different integrals, one could approximate the total expression at once, for example using Leibniz's integral rule. This could lead to better approximations.
- The number of respondents (200) and the number of simulations per response style (100) could be higher. We didn't do this because the computations cost a lot of time. Vectorizing the code will decrease the computation time.

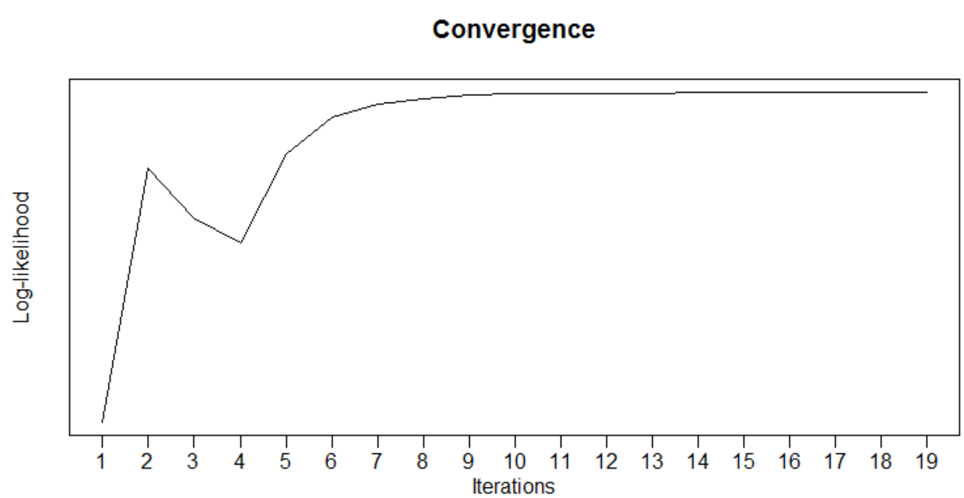

Figure 5: Example of non-monotone convergence

References

[1] J.R. Ashford and R.R. Sowden. Multi-variate probit analysis. *International Biometric Society*, 26(3):535–546, 1970.

[2] Princeton Survey Research Associates. *Pew Internet and American Life Project*, 2003.

[3] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability*. Athena Scientific, 1st edition, 2002.

[4] A. W. Bowman and A. Azzalini. *R package `sm`: nonparametric smoothing methods (version 2.2-5)*. University of Glasgow, UK and Università di Padova, Italia, 2013.

[5] B. Efron and D.V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65(3):457–487, 1978.

[6] A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, 2009.

[7] A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2013. R package version 0.9-9995.

[8] E. Lesaffre and G. Molenberghs. Multivariate probit analysis: A neglected procedure in medical statistics. *Statistics in Medicine*, 10(9):1391–1403, 1991.

[9] R.L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41:351–360, 1954.

[10] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[11] P.C. Schoonees, M. Van de Velden, and P.J.F. Groenen. Constrained dual scaling for detecting response styles in categorical data. *Econometric Institute Report*, 10, 2013.

[12] A. Schwrorer and P. Hovey. Newton-raphson versus fisher scoring algorithms in calculating maximum likelihood estimates. *Electronic Proceedings of Undergraduate Mathematics Day*, 1, 2004.

[13] M. Van de Velden. Detecting response styles by using dual scaling of successive categories. *Econometric Institute Report*, 41, 2007.

[14] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.

Appendix

A Recovering univariate ordered probit parameter estimates

We compare the coefficients of the bivariate ordered probit model and two univariate ordered probit models. The parameter estimates should be the same and we see they almost are. When we increase the number of respondents, the parameter estimates are closer to one another.

parameter	actual value	bivariate model estimate	univariate model estimate
ρ	0	0.11	-
$\beta_{1,1}$	1	1.09	1.09
$\beta_{2,1}$	1	1.09	1.09
$\beta_{1,2}$	0	0.04	0.03
$\beta_{2,2}$	1.5	1.48	1.48
$\alpha_{1,1}$	-1.4	-1.35	-1.35
$\alpha_{2,1}$	-0.4	-0.64	-0.64
$\alpha_{3,1}$	0.6	0.55	0.56
$\alpha_{4,1}$	1.5	1.38	1.39
$\alpha_{1,2}$	-1.4	-1.48	-1.48
$\alpha_{2,2}$	-0.5	-0.44	-0.43
$\alpha_{3,2}$	0.2	0.44	0.45
$\alpha_{4,2}$	1	1.05	1.06

Table 6: Comparison with univariate ordered probit models ($\rho = 0$, $n = 200$)

B Smooth density plots of simulation parameter estimates

For each response style, the parameter estimates of the model are estimated 100 times. The density curves are computed by the R package *sm*, written by Bowman & Azzalini (2013).

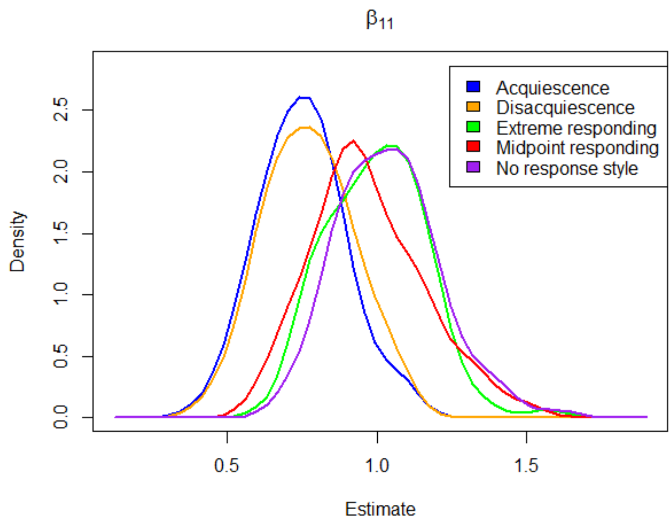

Figure 6: Density plot of parameter estimates of $\beta_{1,1}$ (actual value = 1)

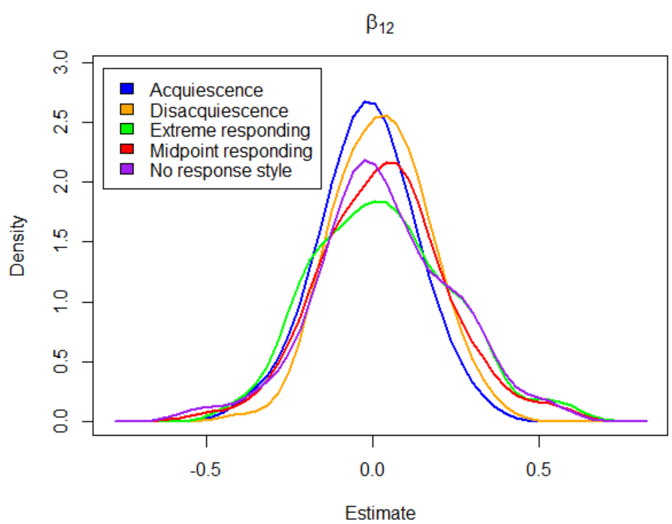

Figure 7: Density plot of parameter estimates of $\beta_{1,2}$ (actual value = 0)

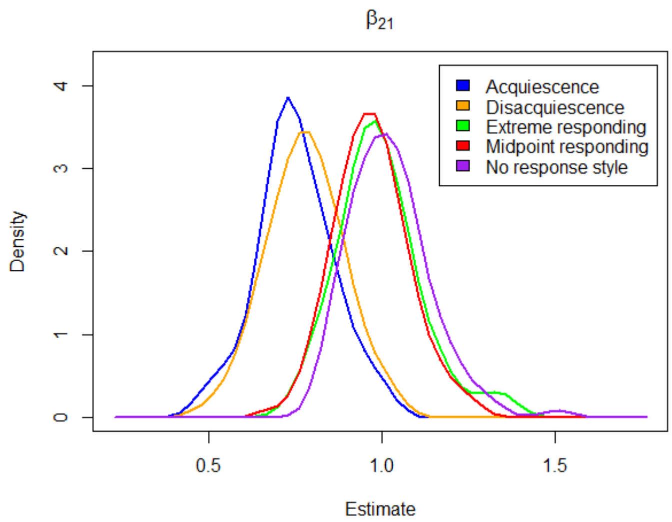

Figure 8: Density plot of parameter estimates of $\beta_{2,1}$ (actual value = 1)

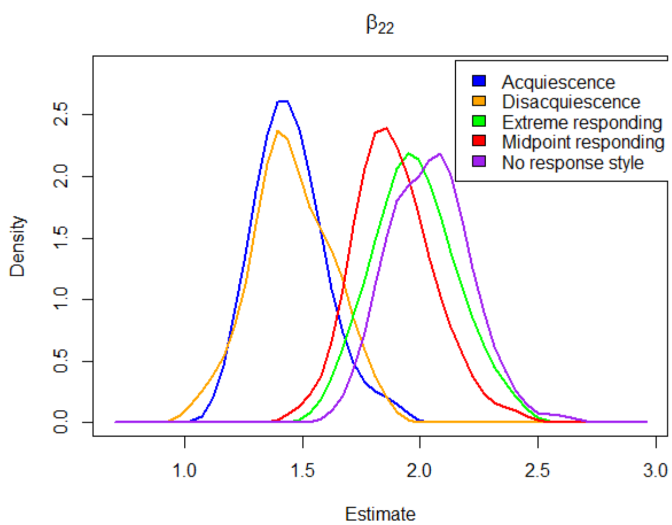

Figure 9: Density plot of parameter estimates of $\beta_{2,2}$ (actual value = 2)

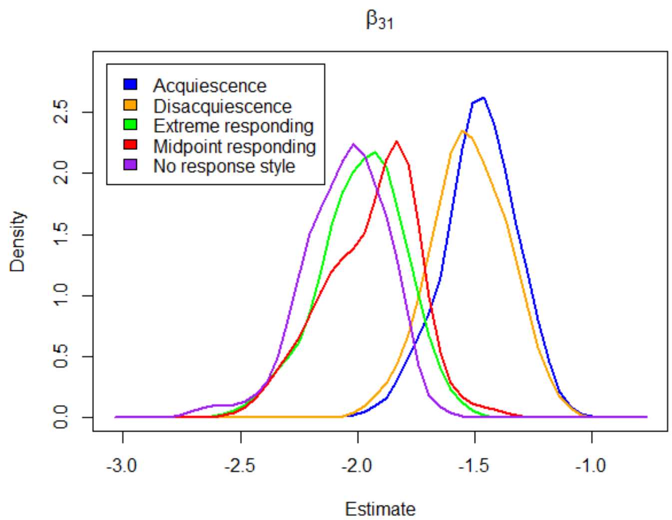

Figure 10: Density plot of parameter estimates of $\beta_{3,1}$ (actual value = -2)

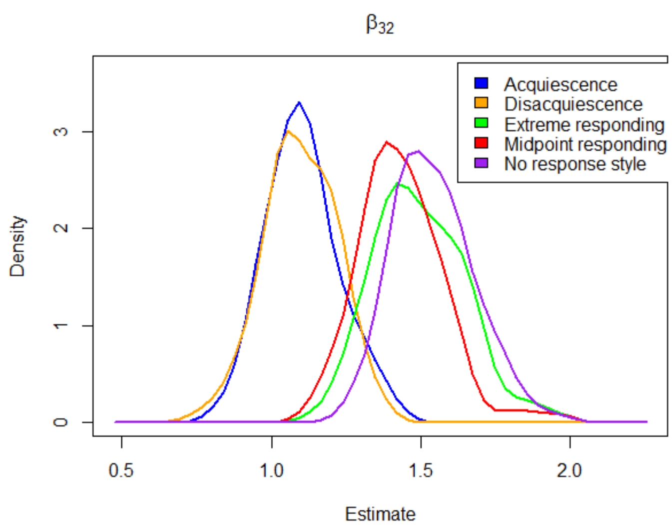

Figure 11: Density plot of parameter estimates of $\beta_{3,2}$ (actual value = 1.5)

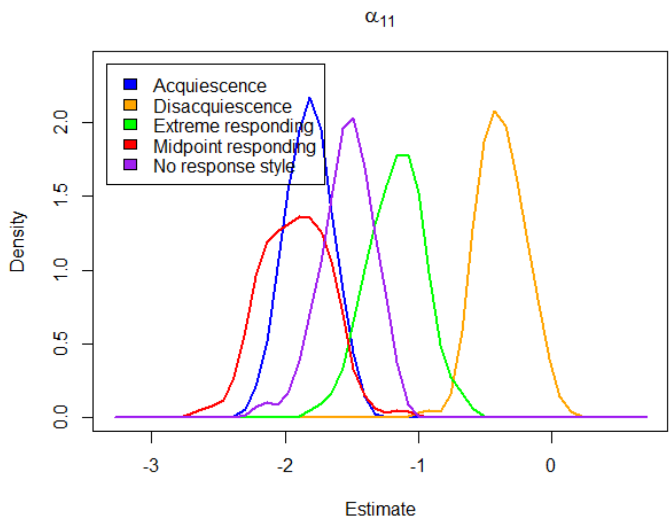

Figure 12: Density plot of parameter estimates of $\alpha_{1,1}$ (actual value = -1.5)

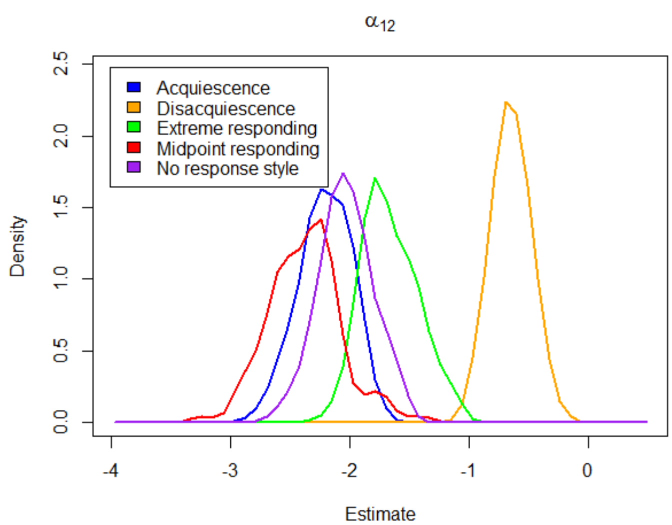

Figure 13: Density plot of parameter estimates of $\alpha_{1,2}$ (actual value = -2)

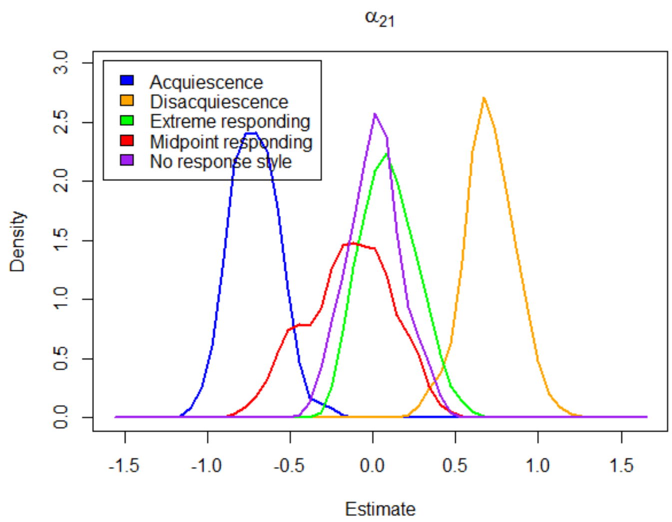

Figure 14: Density plot of parameter estimates of $\alpha_{2,1}$ (actual value = 0)

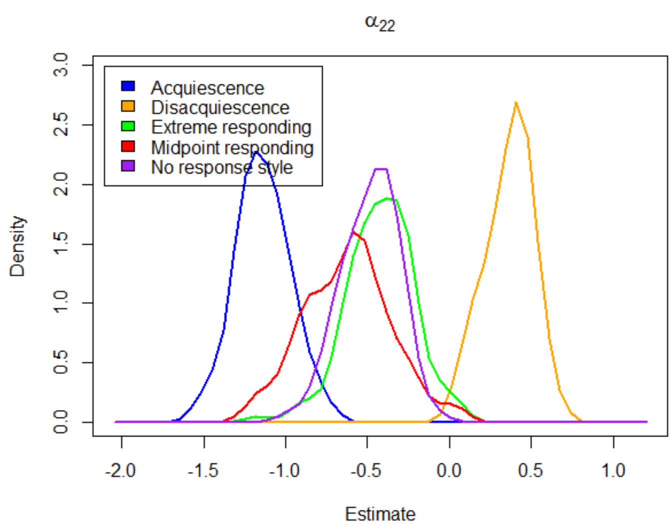

Figure 15: Density plot of parameter estimates of $\alpha_{2,2}$ (actual value = -0.5)

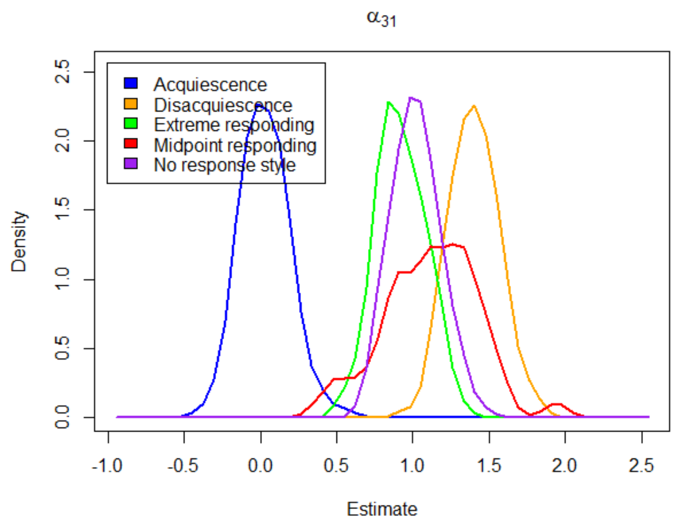

Figure 16: Density plot of parameter estimates of $\alpha_{3,1}$ (actual value = 1)

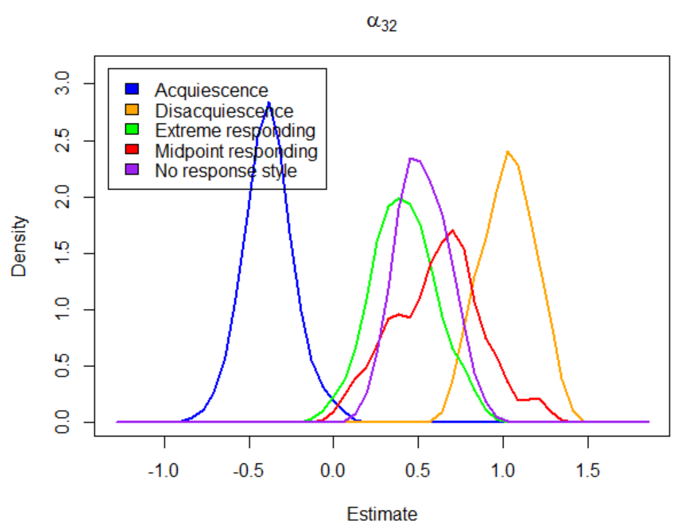

Figure 17: Density plot of parameter estimates of $\alpha_{3,2}$ (actual value = 0.5)

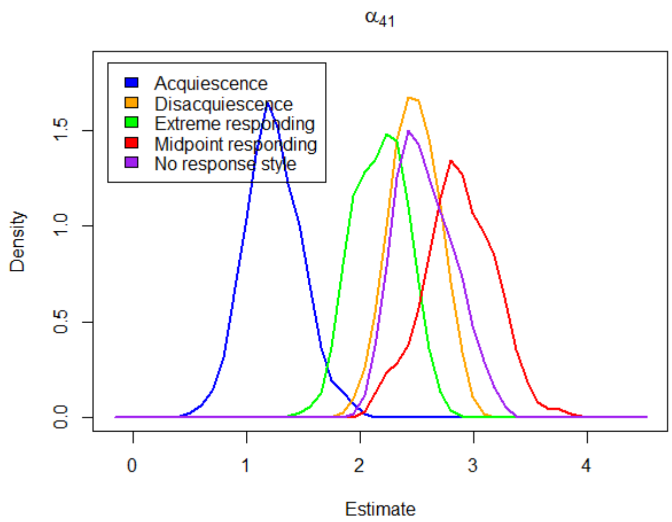

Figure 18: Density plot of parameter estimates of $\alpha_{4,1}$ (actual value = 2.5)

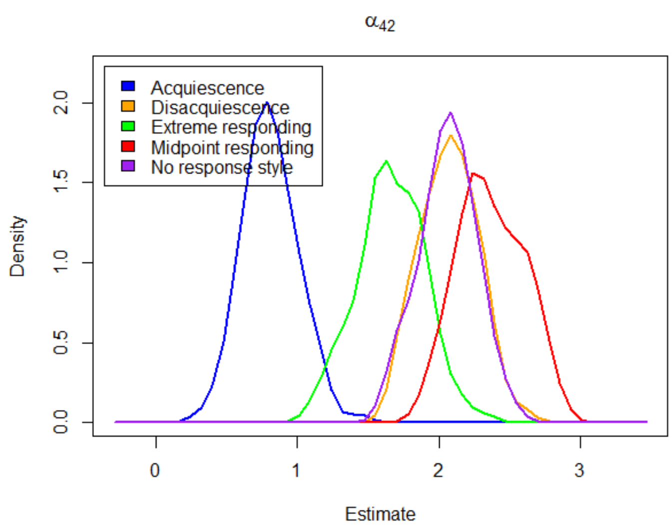

Figure 19: Density plot of parameter estimates of $\alpha_{4,2}$ (actual value = 2)

C Applying (dis)acquiescence to a known group

We have applied the acquiescence and disacquiescence response styles to "a known part" of the respondents. The covariate x_1 is a dummy-variable that indicates whether a respondent belong to this group.

parameter	actual value	no response style estimate	acquiescence estimate	disacquiescence estimate
ρ	0.5	0.51 (0.08)	0.49 (0.10)	0.50 (0.09)
$\beta_{1,1}$	1	1.04 (0.17)	-0.89 (0.28)	2.89 (0.36)
$\beta_{2,1}$	1	1.02 (0.12)	1.01 (0.11)	1.04 (0.12)
$\beta_{3,1}$	-2	-2.05 (0.17)	-2.06 (0.15)	-2.06 (0.19)
$\beta_{1,2}$	0	0.03 (0.20)	-2.07 (0.34)	1.98 (0.33)
$\beta_{2,2}$	2	2.04 (0.17)	2.04 (0.21)	2.05 (0.18)
$\beta_{3,2}$	1.5	1.54 (0.13)	1.54 (0.15)	1.55 (0.15)
$\alpha_{1,1}$	-1.5	-1.54 (0.19)	-3.50 (0.37)	0.46 (0.27)
$\alpha_{2,1}$	0	0.01 (0.15)	-2.00 (0.29)	1.90 (0.32)
$\alpha_{3,1}$	1	1.02 (0.15)	-0.93 (0.26)	2.85 (0.35)
$\alpha_{4,1}$	2.5	2.58 (0.24)	0.74 (0.29)	4.37 (0.47)
$\alpha_{1,2}$	-2	-2.03 (0.22)	-4.13 (0.46)	0.04 (0.26)
$\alpha_{2,2}$	-0.5	-0.48 (0.17)	-2.65 (0.39)	1.51 (0.31)
$\alpha_{3,2}$	0.5	0.53 (0.14)	-1.59 (0.32)	2.46 (0.33)
$\alpha_{4,2}$	2	2.06 (0.19)	0.09 (0.27)	3.87 (0.38)

Table 7: Applying (dis)acquiescence to "a known part" of the respondents ($n = 200$)