Distance models for analysis of multivariate binary data

Worku, H.M.

Appendix A

List of Tables that belong to Chapter 2

Table A.1: Parameter estimates of the 2-way interaction logistic regression model fitted on the nonconvergence data. For simplicity, we denote the design variables as, a: type of indicators; b: number of indicators; c: factor structure; d: correlation between underlying latent variables; and, e: sample size.

Effect		Estimates (S.E.)	p-value	OR
Intercept		-22.14 (2.548)	< 0.0001	
Type of Indicators (a)	BLR	11.27 (1.578)	< 0.0001	78,766.54**
	BMR	5.96 (1.637)	0.0003	385.92**
Number of Indicators (b)	6	6.58 (1.451)	< 0.0001	717.83**
	10	3.27 (1.534)	0.0333	26.20**
Factor structure (c)	Weak	9.72 (1.555)	< 0.0001	16,724.23**
	Moderate	6.53 (1.573)	< 0.0001	683.68**
Correlation	Independence	0.06 (0.437)	0.8878	1.06
between Factors (d)	Moderate	-2.57 (0.729)	0.0004	0.08**
Sample size (e)	50	13.39 (2.438)	< 0.0001	655,792.70**
	100	10.03 (2.44)	< 0.0001	22,781.63**
	300	4.88 (2.471)	0.0481	131.95**
a $\times$ b	[a=BLR] $\times$ [b=6]	-1.70 (0.293)	< 0.0001	0.18**
	[a=BMR] $\times$ [b=6]	-0.34 (0.295)	0.2500	0.71
	[a=BLR] $\times$ [b=10]	-1.32 (0.300)	< 0.0001	0.27*
	[a=BMR] $\times$ [b=10]	-0.66 (0.302)	0.0286	0.52
a $\times$ c	[a=BLR] $\times$ [c=Weak]	-4.50 (0.646)	< 0.0001	0.01**
	[a=BMR] $\times$ [c=Weak]	-2.85 (0.647)	< 0.0001	0.06**
	[a=BLR] $\times$ [c=Moderate]	-3.27 (0.652)	< 0.0001	0.04**
	[a=BMR] $\times$ [c=Moderate]	-2.17 (0.651)	0.0009	0.11**

a × d	[a=BLR] × [d=Independence]	-1.31 (0.204)	< 0.0001	0.27*
	[a=BMR] × [d=Independence]	-1.15 (0.201)	< 0.0001	0.32
	[a=BLR] × [d=Moderate]	-0.72 (0.214)	0.0008	0.49
	[a=BMR] × [d=Moderate]	-0.84 (0.210)	< 0.0001	0.43
a × e	[a=BLR] × [e=50]	-2.51 (1.404)	0.0738	0.08**
	[a=BMR] × [e=50]	0.15 (1.470)	0.9213	1.16
	[a=BLR] × [e=100]	-0.65 (1.408)	0.6441	0.52
	[a=BMR] × [e=100]	1.09 (1.473)	0.4597	2.97
	[a=BLR] × [e=300]	0.85 (1.448)	0.5591	2.33
	[a=BMR] × [e=300]	1.68 (1.512)	0.2669	5.36**
b × c	[b=6] × [c=Weak]	-0.75 (0.149)	< 0.0001	0.47
	[b=10] × [c=Weak]	-0.16 (0.148)	0.2745	0.85
	[b=6] × [c=Moderate]	-0.02 (0.152)	0.9070	0.98
	[b=10] × [c=Moderate]	0.01 (0.150)	0.9258	1.01
b × d	[b=6] × [d=Independence]	1.07 (0.129)	< 0.0001	2.92
	[b=10] × [d=Independence]	0.56 (0.126)	< 0.0001	1.76
	[b=6] × [d=Moderate]	1.12 (0.132)	< 0.0001	3.06*
	[b=10] × [d=Moderate]	0.60 (0.130)	< 0.0001	1.82
b × e	[b=6] × [e=50]	-4.27 (1.413)	0.0025	0.01**
	[b=10] × [e=50]	-1.80 (1.498)	0.2300	0.17**
	[b=6] × [e=100]	-3.21 (1.413)	0.0230	0.04**
	[b=10] × [e=100]	-1.28 (1.497)	0.3927	0.28*
	[b=6] × [e=300]	-2.08 (1.417)	0.1418	0.13**
	[b=10] × [e=300]	-0.74 (1.502)	0.6208	0.48
c × d	[c=Weak] × [d=Independence]	-0.40 (0.131)	0.0024	0.67
	[c=Moderate] × [d=Independence]	0.00 (0.132)	0.9850	1.00
	[c=Weak] × [d=Moderate]	-0.31 (0.133)	0.0181	0.73
	[c=Moderate] × [d=Moderate]	-0.53 (0.133)	< 0.0001	0.59
c × e	[c=Weak] × [e=50]	-4.29 (1.404)	0.0022	0.01**
	[c=Moderate] × [e=50]	-2.79 (1.421)	0.0497	0.06**
	[c=Weak] × [e=100]	-3.37 (1.404)	0.0164	0.03**
	[c=Moderate] × [e=100]	-2.17 (1.421)	0.1259	0.11**
	[c=Weak] × [e=300]	-1.38 (1.410)	0.3288	0.25**
	[c=Moderate] × [e=300]	-0.76 (1.427)	0.5921	0.47
d × e	[d=Independence] × [e=50]	0.62 (0.330)	0.0602	1.86
	[d=Moderate] × [e=50]	2.89 (0.666)	< 0.0001	17.98**
	[d=Independence] × [e=100]	0.77 (0.329)	0.0185	2.17
	[d=Moderate] × [e=100]	2.92 (0.665)	< 0.0001	18.57**
	[d=Independence] × [e=300]	1.12 (0.331)	0.0007	3.06*

	[d=Moderate] $\times$ [e=300]	2.60 (0.666)	< 0.0001	13.40**

*medium effect size, according to Ferguson (2009).

**large effect size, according to Ferguson (2009).

Table A.2: Parameter estimates of the 2-way interaction logistic regression model fitted on the *Heywood* data. For simplicity, we denote the design variables as, a: type of indicators; b: number of indicators; c: factor structure; d: correlation between underlying latent variables; and, e: sample size.

Effect		Estimates (S.E.)	p-value	OR
Intercept		-20.61 (2.813)	< 0.0001	
Type of Indicators (a)	BLR	8.73 (1.455)	< 0.0001	6,179.20
	BMR	0.33 (1.950)	0.8664	1.39
Number of Indicators (b)	6	4.13 (1.456)	0.0045	62.28
	10	-0.49 (1.934)	0.7982	0.61
Factor structure (c)	Weak	6.96 (1.504)	< 0.0001	1,049.37
	Moderate	1.65 (1.962)	0.3992	5.23
Correlation	Independence	5.52 (1.450)	0.0001	248.30
between Factors (d)	Moderate	3.75 (1.571)	0.0170	42.55
Sample size (e)	50	13.01 (2.708)	< 0.0001	445,359.70
	100	7.53 (2.709)	0.0054	1,865.31
	300	1.85 (2.794)	0.5075	6.37
a $\times$ b	[a=BLR] $\times$ [b=6]	-1.46 (0.285)	< 0.0001	0.23^{**}
	[a=BMR] $\times$ [b=6]	0.72 (0.349)	0.0390	2.06
	[a=BLR] $\times$ [b=10]	-0.97(0.300)	0.0013	0.38
	[a=BMR] $\times$ [b=10]	0.50 (0.360)	0.1687	1.64
a $\times$ c	[a=BLR] $\times$ [c=Weak]	-3.14 (0.322)	< 0.0001	0.04^{**}
	[a=BMR] $\times$ [c=Weak]	-0.11 (0.414)	0.7826	0.89
	[a=BLR] $\times$ [c=Moderate]	-2.53 (0.325)	< 0.0001	0.08^{**}
	[a=BMR] $\times$ [c=Moderate]	-0.04 (0.416)	0.9312	0.97
a $\times$ d	[a=BLR] $\times$ [d=Independence]	-1.61 (0.197)	< 0.0001	0.20^{**}
	[a=BMR] $\times$ [d=Independence]	-0.96 (0.194)	< 0.0001	0.38
	[a=BLR] $\times$ [d=Moderate]	-1.25 (0.201)	< 0.0001	0.29^{*}
	[a=BMR] $\times$ [d=Moderate]	-0.78 (0.197)	< 0.0001	0.46
a $\times$ e	[a=BLR] $\times$ [e=50]	-5.23 (1.376)	0.0001	0.01^{**}
	[a=BMR] $\times$ [e=50]	-0.76 (1.861)	0.6843	0.47
	[a=BLR] $\times$ [e=100]	-2.61 (1.373)	0.0578	0.07^{**}
	[a=BMR] $\times$ [e=100]	0.02 (1.861)	0.9928	1.02
	[a=BLR] $\times$ [e=300]	-0.65 (1.380)	0.6355	0.52
	[a=BMR] $\times$ [e=300]	0.51 (1.868)	0.7858	1.66
b $\times$ c	[b=6] $\times$ [c=Weak]	0.08 (0.370)	0.8364	1.08
	[b=10] $\times$ [c=Weak]	1.23 (0.474)	0.0096	3.41^{*}
	[b=6] $\times$ [c=Moderate]	1.44 (0.434)	0.0009	4.22^{**}
	[b=10] $\times$ [c=Moderate]	1.81 (0.530)	0.0006	6.12^{**}
b $\times$ d	[b=6] $\times$ [d=Independence]	-0.14 (0.309)	0.6619	0.87

	[b=10] × [d=Independence]	0.28 (0.329)	0.3979	1.32
	[b=6] × [d=Moderate]	-0.37 (0.311)	0.2353	0.69
	[b=10] × [d=Moderate]	0.01 (0.333)	0.9721	1.01
b × e	[b=6] × [e=50]	-1.49 (1.337)	0.2662	0.23**
	[b=10] × [e=50]	0.54 (1.815)	0.7656	1.72
	[b=6] × [e=100]	0.50 (1.341)	0.7107	1.64
	[b=10] × [e=100]	1.83 (1.817)	0.3138	6.23**
	[b=6] × [e=300]	1.01 (1.363)	0.4581	2.75
	[b=10] × [e=300]	1.98 (1.835)	0.2806	7.24**
c × d	[c=Weak] × [d=Independence]	-1.37 (0.361)	0.0001	0.25**
	[c=Moderate] × [d=Independence]	-0.56 (0.367)	0.1261	0.57
	[c=Weak] × [d=Moderate]	-1.04 (0.373)	0.0051	0.35
	[c=Moderate] × [d=Moderate]	-0.63 (0.381)	0.0978	0.53
c × e	[c=Weak] × [e=50]	-2.55 (1.377)	0.0646	0.08**
	[c=Moderate] × [e=50]	0.26 (1.852)	0.8906	1.29
	[c=Weak] × [e=100]	-0.48 (1.378)	0.7273	0.62
	[c=Moderate] × [e=100]	2.10 (1.853)	0.2577	8.14**
	[c=Weak] × [e=300]	2.02 (1.503)	0.1793	7.52**
	[c=Moderate] × [e=300]	4.06 (1.948)	0.0372	57.89**
d × e	[d=Independence] × [e=50]	-2.78 (1.361)	0.0414	0.06**
	[d=Moderate] × [e=50]	-1.46 (1.486)	0.3263	0.23**
	[d=Independence] × [e=100]	-1.98 (1.358)	0.1456	0.14**
	[d=Moderate] × [e=100]	-0.87 (1.483)	0.5558	0.42
	[d=Independence] × [e=300]	-0.90 (1.364)	0.5115	0.41
	[d=Moderate] × [e=300]	-0.24 (1.489)	0.8705	0.78

*medium effect size, according to Ferguson (2009).

**large effect size, according to Ferguson (2009).

Table A.3: Effect size of the 2-way interaction ANOVA model fitted on the average correlations reported in Table 2.5. The design variables are denoted by letters, i.e., a: type of indicators; b: number of indicators; c: factor structure; d: correlation between latent variables; and e: sample size.

Effect	F (df)	p-value	η^2
Type of Indicators (a)	35549.88 (2)	0.000	0.736**
Number of Indicators (b)	13661.71 (2)	0.000	0.517**
Factor structure (c)	44178.4 (2)	0.000	0.776**
Correlation between Factors (d)	2643.23 (2)	0.000	0.172**
Sample size (e)	641.25 (3)	0.000	0.070*
Interactions			
a × b	270.02 (4)	0.000	0.041
a × c	1704.7 (4)	0.000	0.211**
a × d	7.76 (4)	0.000	0.001
a × e	9.03 (6)	0.000	0.002
b × c	74.34 (4)	0.000	0.012
b × d	4.53 (4)	0.001	0.001
b × e	3.09 (6)	0.005	0.001
c × d	18.07 (4)	0.000	0.003
c × e	46.14 (6)	0.000	0.011
d × e	7.03 (6)	0.000	0.002

*medium effect size, according to Cohen (1988).

**large effect size, according to Cohen (1988).

Table A.4: Observed power for the relationship between X_7 and the second factor, $\gamma_{72} = 0.10$. The number of replications per cell differ because of improper solutions. Dashed lines indicate no valid results were obtained for that cell.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	—	—	—	0.00	0.06	0.00	0.07	0.08	0.13	0.39	0.37	0.57
		Moderate	—	—	—	0.04	0.03	0.04	0.05	0.06	0.11	0.44	0.47	0.47
		Strong	—	—	—	0.00	0.00	0.08	0.02	0.06	0.09	0.40	0.45	0.53
	Moderate	Independence	—	—	—	0.00	0.00	0.00	0.03	0.05	0.05	0.40	0.47	0.59
		Moderate	—	—	—	0.00	0.00	0.00	0.07	0.04	0.06	0.47	0.42	0.56
		Strong	0.00	—	—	0.00	0.06	0.00	0.08	0.07	0.08	0.39	0.52	0.50
	Strong	Independence	—	0.00	0.00	0.00	0.00	0.08	0.12	0.03	0.17	0.40	0.52	0.43
		Moderate	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.06	0.08	0.50	0.49	0.50
		Strong	—	0.00	—	0.00	0.00	0.00	0.00	0.00	0.11	0.54	0.53	0.38
BMR	Weak	Independence	0.00	0.03	0.05	0.07	0.06	0.10	0.05	0.14	0.13	0.60	0.65	0.83
		Moderate	0.06	0.00	0.01	0.05	0.08	0.05	0.09	0.11	0.14	0.66	0.66	0.68
		Strong	0.03	0.07	0.03	0.06	0.08	0.04	0.11	0.12	0.11	0.63	0.66	0.77
	Moderate	Independence	0.00	0.07	0.08	0.10	0.04	0.03	0.12	0.11	0.14	0.67	0.70	0.84
		Moderate	0.00	0.06	0.08	0.05	0.07	0.09	0.08	0.15	0.11	0.67	0.75	0.68
		Strong	0.06	0.07	0.03	0.04	0.12	0.07	0.13	0.16	0.14	0.67	0.78	0.82
	Strong	Independence	0.00	0.00	0.00	0.14	0.00	0.13	0.14	0.13	0.18	0.60	0.64	0.70
		Moderate	0.00	0.00	—	0.09	0.00	0.00	0.00	0.21	0.33	0.62	0.75	0.70
		Strong	0.00	—	0.00	0.00	0.25	0.25	0.00	0.00	0.00	0.62	0.71	0.73
Interval	Weak	Independence	0.05	0.12	0.05	0.12	0.10	0.14	0.12	0.06	0.06	0.65	0.80	0.81
		Moderate	0.10	0.08	0.05	0.09	0.11	0.12	0.11	0.12	0.13	0.68	0.79	0.78
		Strong	0.09	0.13	0.06	0.07	0.15	0.13	0.11	0.18	0.14	0.63	0.81	0.87
	Moderate	Independence	0.09	0.08	0.06	0.10	0.14	0.11	0.13	0.15	0.11	0.82	0.80	0.76
		Moderate	0.08	0.07	0.08	0.07	0.11	0.05	0.10	0.16	0.11	0.67	0.81	0.86
		Strong	0.13	0.12	0.15	0.08	0.10	0.15	0.14	0.18	0.11	0.79	0.88	0.87
	Strong	Independence	0.13	0.14	0.15	0.05	0.13	0.12	0.21	0.17	0.15	0.89	0.84	0.88
		Moderate	0.14	0.09	0.06	0.06	0.11	0.06	0.21	0.17	0.16	0.80	0.94	0.89
		Strong	0.11	0.18	0.11	0.12	0.09	0.12	0.13	0.13	0.13	0.84	0.89	0.85

Table A.5: Observed power for the relationship between X_4 and the second factor, $\gamma_{42} = 0.95$. The number of replications per cell differ because of improper solutions. Dashed lines indicate no valid results were obtained for that cell.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	—	—	—	0.38	0.43	0.62	0.97	1.00	1.00	1.00	1.00	1.00
		Moderate	—	—	—	0.38	0.53	0.64	0.98	1.00	1.00	1.00	1.00	1.00
		Strong	—	—	—	0.17	0.43	0.72	1.00	1.00	1.00	1.00	1.00	1.00
	Moderate	Independence	—	—	—	0.31	0.50	0.83	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	—	—	—	0.53	0.70	0.80	0.99	1.00	1.00	1.00	1.00	1.00
		Strong	0.00	—	—	0.26	0.56	0.75	1.00	1.00	1.00	1.00	1.00	1.00
	Strong	Independence	—	0.00	0.00	0.00	0.70	0.58	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.00	0.00	0.00	0.80	0.62	0.36	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	—	0.00	—	1.00	0.33	0.50	0.93	1.00	1.00	1.00	1.00	1.00
BMR	Weak	Independence	0.71	0.67	0.70	0.92	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.50	0.69	0.79	0.89	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	0.74	0.72	0.79	0.90	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	Moderate	Independence	0.64	0.72	0.88	0.97	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.55	0.64	0.75	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	0.55	0.75	0.84	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Strong	Independence	1.00	0.67	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.50	0.33	—	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	0.50	—	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Interval	Weak	Independence	0.67	0.82	0.89	0.90	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.53	0.79	0.85	0.97	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	0.58	0.81	0.85	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Moderate	Independence	0.77	0.91	0.89	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.82	0.89	0.93	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	0.80	0.88	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Strong	Independence	0.93	0.92	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Moderate	0.90	0.97	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Strong	0.91	0.94	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Appendix B

Identification of the BIPC Model in Chapter 3

In this Section, we address identification issues of the BIPC model. As shown in equation (3.25), the marginal models are represented in the BIPC model as,

$$\pi_{i1\bullet} = \frac{\exp(-0.5\delta_{i1\bullet})}{\exp(-0.5\delta_{i0\bullet}) + \exp(-0.5\delta_{i1\bullet})},$$

$$\pi_{i\bullet 1} = \frac{\exp(-0.5\delta_{i\bullet 1})}{\exp(-0.5\delta_{i\bullet 0}) + \exp(-0.5\delta_{i\bullet 0})},$$

where $\delta_{il\bullet} = \sum_{m=1}^{2} (\eta_{im} - \gamma_{l\bullet m})^2$ and $\delta_{i\bullet l} = \sum_{m=1}^{2} (\eta_{im} - \gamma_{\bullet lm})^2$, $l = 0, 1$. Like simple logistic regression (LR) model, the BIPC model can also be represented using log-odds, i.e.,

$$\log\left[\frac{\pi_{i1\bullet}}{1-\pi_{i1\bullet}}\right] = 0.5\delta_{i0\bullet} - 0.5\delta_{i1\bullet},$$

$$\log\left[\frac{\pi_{i\bullet 1}}{1-\pi_{i\bullet 1}}\right] = 0.5\delta_{i\bullet 0} - 0.5\delta_{i\bullet 1}.$$

In the BIPC model, each binary response variable is positioned on one and only dimension, and thus the class coordinates are specified as follows:

$$\gamma_{l\bullet} = \begin{bmatrix} \gamma_{0\bullet 1} & 0 \\ \gamma_{1\bullet 1} & 0 \end{bmatrix} \text{ and } \quad \gamma_{\bullet l} = \begin{bmatrix} 0 & \gamma_{\bullet 02} \\ 0 & \gamma_{\bullet 12} \end{bmatrix},$$

where the first binary response (Y_{i1}) is positioned on the first dimension and the second binary response variable (Y_{i2}) on the second dimension.

Suppose x_{i1} represents one of the predictor variables. Let us now simplify the log-odds model by replacing the above class points. That is,

$$\begin{aligned}
\log\left[\frac{\pi_{i1\cdot}}{1-\pi_{i1\cdot}}\right] &= \sum_{m=1}^{2}(\eta_{im}\gamma_{1\cdot m}) - \sum_{m=1}^{2}(\eta_{im}\gamma_{0\cdot m}) + 0.5\sum_{m=1}^{2}\gamma_{0\cdot m}^2 - 0.5\sum_{m=1}^{2}\gamma_{1\cdot m}^2 \\
&= \eta_{i1}(\gamma_{1\cdot 1}-\gamma_{0\cdot 1}) + 0.5\times(\gamma_{0\cdot 1}^2-\gamma_{1\cdot 1}^2) \\
&= (\beta_{01}+\beta_{11}x_{i1})(\gamma_{1\cdot 1}-\gamma_{0\cdot 1}) + 0.5\times(\gamma_{0\cdot 1}^2-\gamma_{1\cdot 1}^2) \\
&= \beta_{01}(\gamma_{1\cdot 1}-\gamma_{0\cdot 1}) + \beta_{11}(\gamma_{1\cdot 1}-\gamma_{0\cdot 1})x_{i1} + 0.5\times(\gamma_{0\cdot 1}^2-\gamma_{1\cdot 1}^2) \\
&= \beta_{01}^* + \beta_{11}^* x_{i1},
\end{aligned}$$

where $\beta_{01}^* = \beta_{01}(\gamma_{1\cdot 1}-\gamma_{0\cdot 1}) + 0.5\times(\gamma_{0\cdot 1}^2-\gamma_{1\cdot 1}^2)$ and $\beta_{11}^* = \beta_{11}(\gamma_{1\cdot 1}-\gamma_{0\cdot 1})$. So, for a unit increase in x_{i1} the log-odds in the BIPC model changes by $\beta_{11}(\gamma_{1\cdot 1}-\gamma_{0\cdot 1})$. Similarly, the simplified log-odds form of the second binary response variable becomes $\log[\pi_{i\cdot 1}/(1-\pi_{i\cdot 1})] = \beta_{02}(\gamma_{\cdot 12}-\gamma_{\cdot 02}) + \beta_{12}(\gamma_{\cdot 12}-\gamma_{\cdot 02})x_{i1} + 0.5\times(\gamma_{\cdot 02}^2-\gamma_{\cdot 12}^2) = \beta_{02}^* + \beta_{12}^* x_{i1}$.

At this stage, the BIPC model is not identified since both the regression weights and the class coordinates influence the distance model parameters. For unique identification of model parameters, we must impose restrictions on the class coordinates. This can be achieved, for example, by setting a unit difference between the class coordinates. That is, $\gamma_{1\cdot 1} = 1$ and $\gamma_{0\cdot 1} = 0$ for the first response variable, Y_{i1}; and $\gamma_{\cdot 12} = 1$ and $\gamma_{\cdot 02} = 0$ for Y_{i2}. Thus, $\beta_{1s}^* = \beta_{1s}$ and $\beta_{0s}^* = \beta_{0s} - 0.5$, where $s = 1, 2$.

Appendix C

Chapter 4: Exploring latent variable models with few number of indicators per factor

Recently, in clinical psychological research factor analysis and structural equation models have been proposed for the analysis of comorbidity of depressive and anxiety disorders (Krueger, 1999; Beesdo-Baum et al., 2009). A typical characteristic of these models is that the indicators are dichotomous, i.e whether someone has or does not have a particular disorder, and that there are only a few indicators per latent variable, i.e. 2 or 3. These authors found for depressive and anxiety disorders two underlying factors: 'fear' with indicators social phobia (SP), and panic disorder (PD), and 'distress' with indicators major depressive disorder (MDD), dysthmia (DYST), and generalized anxiety disorder (GAD).

We tried to replicate these findings using data from the Netherlands Study for Depression and Anxiety (NESDA, Penninx et al., 2008). In this study, measurements were obtained on N=2,938 subjects of age between $18 - 65$ years with an average age of 42 (S.D. $= 13.1$), 66.5% were female, and the average number of years of education attained was 12.2 (S.D. $= 3.3$). In the NESDA study measurements are obtained on each of the five disorders, i.e., about 37.1% of the subjects in the study had MDD, 10.2% had DYST,

Table C.1: Fit statistics for the factor models fitted on the NESDA data.

Model	# parameters	RMSEA	CFI
1-factor	10	0.082	0.956
2-factor (fear-distress)	11	0.034	0.994
2-factor (anxiety-depression)	11	0.071	0.974

15.3% had GAD, 22.4% had SP, and 28.6% had PD.

With these data it is possible to investigate the interplay between personality traits and co-morbidity of mental disorders. We considered Big-Five personality traits, i.e., Neuroticism (N), Extraversion (E), Openness to experience (O), Agreeableness (A), and Conscientiousness (C). After establishing the measurement model we would like to see the influence of the personality traits and the background variables (age, gender, and education) in explaining the latent variables.

The first step in the analysis was to establish the measurement model using a Confirmatory Factor Analysis (CFA). Besides the fear-distress theory there are two other competing theories (Krueger, 1999; Beesdo-Baum et al., 2009; Spinhoven et al., 2013): the first is the anxiety-depression theory which differs from the distress-fear with respect to generalized anxiety disorder. The anxiety-depression theory places GAD on the other latent factor. Finally, there is a single factor theory where all five disorders are indicators of a single underlying factor.

In Table C.1 we present fit statistics of the three measurement models applied to the NESDA data. The 2-factor model with fear and distress factors fitted the NESDA data best with fit statistics of RMSEA=0.034 and CFI=0.994. In this selected model, there are a total of eleven (11) parameters: three (3) loadings on the first factor, two (2) loadings on the second factor, a covariance between the factors, and a threshold for each of the five manifest variables.

Table C.2: Parameter estimates with the corresponding standard errors (S.E.) presented in parenthesis for the final 2-factor (fear-distress) model.

Effect	Parameter	Estimate (S.E.)
Fear (F1)		
Gender	γ_{11}	$-0.127(0.072)$
Education†	γ_{21}	$-0.113(0.035)$
Age	γ_{31}	$0.012(0.036)$
Neuroticism†	γ_{41}	$0.985(0.062)$
Extraversion†	γ_{51}	$-0.289(0.045)$
Agreeableness	γ_{61}	$-0.026(0.035)$
Conscientiousness	γ_{71}	$-0.039(0.038)$
Openness	γ_{81}	$-0.014(0.035)$
Distress (F2)		
Gender	γ_{12}	$0.003(0.079)$
Education†	γ_{22}	$-0.166(0.039)$
Age	γ_{32}	$-0.020(0.038)$
Neuroticism†	γ_{42}	$0.954(0.069)$
Extraversion†	γ_{52}	$-0.245(0.047)$
Agreeableness	γ_{62}	$0.002(0.037)$
Conscientiousness†	γ_{72}	$0.112(0.041)$
Openness	γ_{82}	$0.002(0.038)$
Covariance for Factors		
Var(F1)	ψ_{11}	$1.000(--)$
Var(F2)	ψ_{22}	$1.000(--)$
Cov(F1, F2)	ψ_{12}	$0.333(0.053)$
Threshold		
MDD	τ_1	$0.417(0.048)$
DYST	τ_2	$1.533(0.070)$
GAD	τ_3	$1.181(0.058)$
SP	τ_4	$0.943(0.053)$
PD	τ_5	$0.696(0.047)$

† statistically significant effect, i.e., $p < 0.05$.

In the second step, researchers would like to see the effect of the personality variables on the two latent factors. Suppose they would fit a Multiple Indicators and MultIple Causes (MIMIC) model for this purpose. Table C.2 shows the results. From the personality traits, only neuroticism and extraversion had statistically significant effects on both the fear and distress factor. The positive association between neuroticism and both dimensions implies that on average a subject having a higher score on neuroticism would score high on both fear and distress. In the case of extraversion, the association was negative implying a subject with lower score on extraversion would score higher on both fear and distress. Conscientiousness had an effect only on distress. From the background variables, only education had a strong negative association with both factors.

So far, the theory about comorbidity and the influence of personality variables on the latent factors replicates earlier findings. However, when computing and plotting the factor scores some notable results were found. The distribution of the factor scores from the factor analysis are shown in the left column of Figure C.1, and it can be seen that the distribution deviates strongly from normality. It is clear that the assumption of a normally distributed latent variable does not hold. Furthermore, in the right hand side the factor scores are shown for the two latent variables in the MIMIC model. Surprisingly, the distribution of these factor scores seems completely different from the distribution found in the factor analysis.

These surprising results cause doubts about the robustness of latent variable models with just a few dichotomous indicators per factor. However, we did not find large scale simulation studies for such models. The aim of Chapter 2 is to fill this gap. In Chapter 4 a Multivariate Logistic Distance (MLD) model is proposed to analyze multivariate binary data.

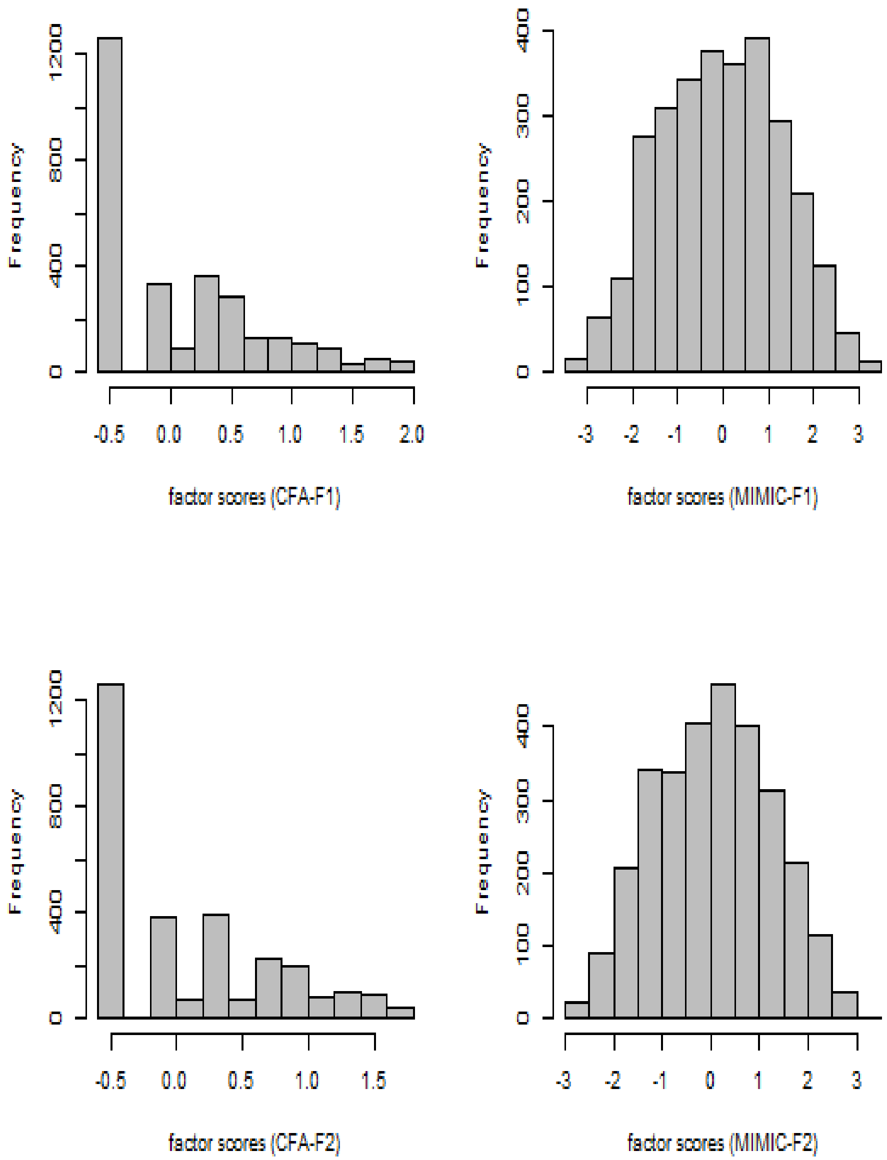

Figure C.1: The distribution of estimated factor scores obtained from the final 2-factor (fear-distress) model. The top panel representing the distribution of scores from the first factor (F1) before and after the inclusion of external variables, respectively; and, the bottom panel for those scores from the second factor (F2) before and after the inclusion of the external variables, respectively.

Appendix D

Chapter 4: Simplification of log-odds representation of the MLD model, Eq. 4.8

Let us simplify log-odds representation of the MLD model (Eq. 4.8) using distress-fear model, where the first dimension (distress) was represented by Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), and Dysthimia (DYST); and the second dimension (fear) was represented by Panic Disorder (PD) and Social Phobia (SP) (Spinhoven et al., 2013). Suppose class coordinates of GAD on the first dimension and PD on the second dimension are set to fixed for identification reasons. Thus, class coordinates of the other response variables become

$$\boldsymbol{\gamma}_2^{\mathsf{GAD}} = \begin{bmatrix} \gamma_{02,1} & 0 \\ \gamma_{12,1} & 0 \end{bmatrix}, \quad \boldsymbol{\gamma}_3^{\mathsf{DYST}} = \begin{bmatrix} \gamma_{03,1} & 0 \\ \gamma_{13,1} & 0 \end{bmatrix}, \quad \boldsymbol{\gamma}_5^{\mathsf{SP}} = \begin{bmatrix} 0 & \gamma_{05,2} \\ 0 & \gamma_{15,2} \end{bmatrix}, \tag{D.1}$$

For demonstration purpose let us focus on GAD response variable of the first dimension and SP of the second dimension, and derive their corresponding simplified version of Eq. 4.8.

The log-odds representation of the multivariate distance model for GAD becomes,

$$\begin{aligned}
\log\left[\frac{\pi_2(\mathbf{x}_i)}{1-\pi_2(\mathbf{x}_i)}\right] &= \sum_{m=1}^{M}\Big\{\beta_{0m}(\gamma_{12,m}-\gamma_{02,m})+0.5(\gamma_{02,m}^2-\gamma_{12,m}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_m(\gamma_{12,m}-\gamma_{02,m})\Big\} \\
&=\Big\{\big\{\beta_{01}(\gamma_{12,1}-\gamma_{02,1})+0.5(\gamma_{02,1}^2-\gamma_{12,1}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_1(\gamma_{12,1}-\gamma_{02,1})\big\} \\
&+\big\{\beta_{02}(\gamma_{12,2}-\gamma_{02,2})+0.5(\gamma_{02,2}^2-\gamma_{12,2}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2(\gamma_{12,2}-\gamma_{02,2})\big\}\Big\} \\
&=\beta_{01}(\gamma_{12,1}-\gamma_{02,1})+0.5(\gamma_{02,1}^2-\gamma_{12,1}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_1(\gamma_{12,1}-\gamma_{02,1}).
\end{aligned}$$

Similarly, the log-odds representation of the multivariate distance model for SP becomes

$$\begin{aligned}
\log\left[\frac{\pi_5(\mathbf{x}_i)}{1-\pi_5(\mathbf{x}_i)}\right] &= \sum_{m=1}^{M}\Big\{\beta_{0m}(\gamma_{15,m}-\gamma_{05,m})+0.5(\gamma_{05,m}^2-\gamma_{15,m}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_m(\gamma_{15,m}-\gamma_{05,m})\Big\} \\
&=\Big\{\big\{\beta_{01}(\gamma_{15,1}-\gamma_{05,1})+0.5(\gamma_{05,1}^2-\gamma_{15,1}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_1(\gamma_{15,1}-\gamma_{05,1})\big\} \\
&+\big\{\beta_{02}(\gamma_{15,2}-\gamma_{05,2})+0.5(\gamma_{05,2}^2-\gamma_{15,2}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2(\gamma_{15,2}-\gamma_{05,2})\big\}\Big\} \\
&=\beta_{02}(\gamma_{15,2}-\gamma_{05,2})+0.5(\gamma_{05,2}^2-\gamma_{15,2}^2) \\
&+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2(\gamma_{15,2}-\gamma_{05,2}).
\end{aligned}$$

Bibliography

Acito, F., & Anderson, R. (1986). A simulation study of factor score indeterminacy. *Journal of Marketing Research*, *23*, 111-118.

Agresti, A. (2002). *Categorical data analysis* (Second ed.). New York: John Wiley and Sons.

Agresti, A. (2007). *An introduction to categorical data analysis* (Second ed.). New Jersey: John Wiley and Sons.

Agresti, A. (2013). *Categorical data analysis* (Third ed.). New York: John Wiley and Sons.

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119-126.

Aitchison, J., & Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restrains. *Annals of Mathematical Statistics*, *29*, 813-828.

Aitchison, J., & Silvey, S. D. (1960). Maximum-likelihood estimation procedures and associated tests of significance. *Journal of the Royal Statistical Society, Series B*, *22*, 154-171.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the second international symposium on information theory* (p. 267-281). Budapest: Akademiai Kiado.

Anderson, J. A. (1984). Regression and order categorical variables. *Journal of Royal Statistics Society B*, *46*(1), 1-30.

Anderson, J. C., & Gerbing, D. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155-173.

Asar, Ö., & Ilk, Ö. (2013). **mmm**: An R package for analyzing multivariate longitudinal data with multivariate marginal models. *Computer Methods and Programs in Biomedicine*, *112*, 649-654.

Ashford, J. R., Morgan, D. C., Rae, S., & Sowden, R. R. (1970). Respiratory symptoms in british coal miners. *The American Review of Respiratory Disease*, *102*(3), 370-81.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In H. Solomon (Ed.), *Studies in item analysis and prediction* (p. 158-176). Stanford, California: Stanford University Press.

Bartholomew, D., & Knott, M. (1999). *Latent variable models and factor analysis.* (2nd ed.). London: Arnold.

Bartholomew, D., Steele, F., Moustaki, I., & Galbraith, J. (2002). *The analysis and interpretation of multivariate data for social scientists.* London: Chapman & Hall.

Bartolucci, F., Colombi, R., & Forcina, A. (2007). An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statistica Sinica*, *17*, 691-711.

Bartolucci, F., Forcina, A., & Dardanoni, V. (2001). Positive quadrant dependence and marginal modelling in two-way tables with ordered margins. *Journal of the American Statistical Association*, *96*(456), 1497-1505.

Beesdo-Baum, K., Höfler, M., Gloster, A. T., Klotsche, J., Lieb, R., Beauducel, A., ... Wittchen, H. U. (2009). The structure of common mental disorders: a replication study in a community sample of adolescents and young adults. *International Journal of Methods in Psychiatric Research*, *18*, 204-220.

Bergsma, W. P. (1997). *Marginal models for categorical variables* (Unpublished doctoral dissertation). Tilburg University.

Bergsma, W. P., Croon, M. A., & Hagenaars, J. A. (2009). *Marginal models for dependent, clustered and longitudinal categorical data.* New York: Springer.

Bergsma, W. P., & Rudas, T. (2002). Marginal models for categorical data. *The Annals of Statistics*, *30*, 140-159.

Bhuyan, M. J., Islam, M. A., & Rahman, M. S. (2018). A bivariate bernoulli model for analyzing malnutrition data. *Health Services and Outcomes Research Methodology*, 1-19. (`https://doi.org/10.1007/s10742-018-0180-9`)

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika*, *35*, 179-197.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*, 3-27.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review Psychology*, *53*, 605-634.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality.* Amsterdam: Sociometric Research Foundation. (Doctoral dissertation, University of Groningen, The Netherlands)

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*(2), 229-242.

Boomsma, A. (2013). Reporting monte carlo studies in structural equation modeling. *Structural Equation Modeling*, *20*, 518-540.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future.* Lincolnwood: Scientific Software International.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (Second ed.). New York: Springer.

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research* *and Perspectives*, *6*(1-2), 25-53.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*(2), 101-133.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford.

Bull, S. B. (1998). Regression models for multiple outcomes in large epidemiological studies. *Statistics in Medicine*, *17*, 2179-2197.

Carey, V., Zeger, S. L., & Diggle, P. (1993). Modeling multivariate binary data with alternating logistic regressions. *Biometrika*, *80*, 517-526.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research*, *29*(4), 468-508.

Cheng, G., Yu, Z., & Huang, J. Z. (2013). The cluster bootstrap consistency in generalized estimating equations. *Journal of Multivariate Analysis*, *115*, 33-47.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*(1), 5-32.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). New York: Academic Press.

Colombi, R., & Forcina, A. (2001). Marginal regression models for the analysis positive association of ordinal response variables. *Biometrika*, *88*(4), 1007-1019.

Coombs, C. H. (1976). *A Theory of Data*. Michigan: Mathesis Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PR-I) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cox, D. R. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *21*(2), 113-120.

Davison, M. L. (1983). *Multidimensional scaling*. USA: John Wiley & Sons, Inc.

De Leeuw, J. (2005). Multidimensional unfolding. In B. S. Everitt & D. C. Howelll (Eds.), *Encyclopedia of statistics in behavioral science* (p. 1289-1294). Wiley.

De Leeuw, J. (2006). *Pseudo-voronoi diagrams for multicategory exponential representations*. UCLA: Statistics Preprints # 463.

De Rooij, M. (2009a). Ideal point discriminant analysis with a special emphasis on visualization. *Psychometrika*, *74*, 317-330.

De Rooij, M. (2009b). Trend vector models for the analysis of change in continuous time for multiple groups. *Computational statistical data analysis*, *53*, 3209-3216.

De Rooij, M. (2011). Transitional ideal point models for longitudinal multinomial outcomes. *Statistical Modelling*, *11*(2), 115-135.

De Rooij, M., & Heiser, W. J. (2005, March). Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. *Psychometrika*, *70*(1), 99-122.

De Rooij, M., & Schouteden, M. (2009). *Mixed effect ideal point models for longitudinal multinomial data.* (Submitted paper)

De Rooij, M., & Worku, H. M. (2012). A warning concerning the estimation of multinomial logistic models with correlated responses in SAS. *Computer Methods and Programs in Biomedicine*, *107*(2), 341-346.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions om structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(2), 119-143.

Elffers, H., Bethlehem, J., & Gill, R. (1978). Indeterminacy problems and the interpretation of factor analysis results. *Statistica Neerlandica*, *32*, 181-199.

Elliot, D. S., Huizinga, D., & Menard, S. (1989). *Multiple problem youth: delinquency, substance use, and mental health problems*. New York: Springer-Verlag.

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532-538.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*, 27-38.

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2008). *Longitudinal data analysis*. London: Chapman and Hall.

Gabriel, K. R. (1971). The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, *58*, 453-467.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: John Wiley & Sons.

Glöckner-Rist, A., & Hoijtink, H. (2003). The best of both words: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling*, *10*(4), 544-565.

Glonek, G. F. V. (1996). A class of regression models for multivariate categorical responses. *Biometrika*, *83*(1), 15-28.

Glonek, G. F. V., & McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B*, *57*, 533-546.

Gower, J. C., & Hand, D. J. (1996). *Biplots*. London: Chapman and Hall.

Gower, J. C., Lubbe, S., & Le Roux, N. (2011). *Understanding biplots*. Chichester: John Wiley & Sons Ltd.

Green, B. F. (1976). On the factor score controversity. *Psychometrika*, *41*(2), 263-266.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430-450.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *British Journal of Statistical Psychology*, *8*(65).

Halekoh, U., Hojsgaard, S., & Yan, J. (2006). The `R` package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, *15*(2).

Hallquist, M. (2012). **MplusAutomation**: Automating **Mplus** model estimation and interpretation [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=MplusAutomation` (R package version 0.5-1)

Heermann, E. F. (1964). The geometry of factorial indeterminacy. *Psychometrika*, *29*(4).

Heermann, E. F. (1966). The algebra of factorial indeterminacy. *Psychometrika*, *31*(4).

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*, 2409-2419.

Heiser, W. J. (1981). *Unfolding analysis of proximity data* (Unpublished doctoral dissertation). Leiden University.

Heiser, W. J. (1987). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (p. 189-221). Springer Verlag.

Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., ... Satariano, W. A. (2010). To GEE or Not to GEE: Comparing population averaged and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, *21*(4), 467-474.

IBM SPSS. (2012). *IBM SPSS statistics version 21.* Boston, Massachusetts: International Business Machines Corp.

Jöreskog, J. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631-639.

Jöreskog, J. G., & Sörbom, D. (1981). *LISREL: Analysis of linear structural relationships by the method of maximum likelihood (version v).* Chicago: National Educational Resources.

Kenneth, A. (1989). *Structural equations with latent variables*. Canada: John Wiley & Sons, Inc.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*(3), 333-351.

Kirk, D. B. (1973). On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika*, *38*(2), 259-268.

Knol, D. L., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*(3), 457-477.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, *56*, 921-926.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Sage publications.

Lang, J. B. (1996). Maximum likelihood methods for a generalized class of log-linear models. *The Annals of Statistics*, *24*(2), 726-752.

Lang, J. B., & Agresti, A. (1994). Simultaneously modelling the joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, *89*, 625-632.

Lawley, D. (1944). The factor analysis of multiple item tests. *Proceedings of the Royal Society of Edinburgh*, *62-A*, 74-82.

Ledermann, W. (1938). The orthogonal transformation of a factorial matrix into itself. *Psychometrika*, *3*(181).

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

Liang, K. Y., Zeger, S. L., & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B (Methodological)*, *54*(1), 3-40.

Lipsitz, S. R., Kim, K., & Zhao, L. P. (1994). Analysis of repeated categorical data using Generalized Estimating Equations. *Statistics in Medicine*, *14*, 1149-1163.

Lipsitz, S. R., Laird, N. M., & Harrington, D. P. (1990). Maximum likelihood regression methods for paired binary data. *Statistics in medicine*, *9*, 1517-1525.

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

MacCallum, R., & Austin, J. (2000). Applications of Structural Equation Modeling in psychological research. *Annual Review Psychology*, *51*, 201-226.

MacLean, R. R., Sofuoglu, M., & Rosenheck, R. (2018). Tobacco and alcohol use disorders: Evaluating multimorbidity. *Addictive Behaviors*, *78*, 59-66.

Mardia, K. V. (1967). Some contributions to the contingency-type bivariate distributions. *Biometrika*, *54*, 235-249.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391-410.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? the number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, *33*(2), 181-220.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Mislevy, R. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, *11*(1), 3-31.

Molenberghs, G., & Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association*, *89*, 633-644.

Molenberghs, G., & Lesaffre, E. (1999). Marginal modeling of multivariate categorical data. *Statistics in Medicine*, *18*, 2237-2255.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York: Springer.

Muthen, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*(4), 551-560.

Muthen, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 43-65.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115-132.

Muthen, L., & Muthen, B. (1998-2012). *Mplus user's guide.* (3rd ed.). Los Angeles, CA: Muthen & Muthen.

Ottenbacher, K. (1991). Interpretation of interaction in factorial analysis of variance design. *Statistics in Medicine*, *10*, 1565-1571.

Palmgren, J. (1989). *Regression Models for Bivariate Binary Responses* (Working Paper No. 101). Seatle, WA: University of Washington.

Pan, W. (2001). Akaike's information criterion in Generalized Estimating Equations. *Biometrics*, *57*, 120-125.

Park, T. (1994). Multivariate regression models for discrete and continuous repeated measurements. *Communications in Statistics - Theory and Methods*, *23*, 1547-1564.

Paxton, P., Curran, P., Bollen, K., Kirby, J., & Fen, C. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *35*(8), 287-312.

Pearson, K. (1900). On the correlation of characters not quantitatively measurable. *Royal Society of Philosophical Transactions, Series A*, *195*, 1-47.

Penninx, B. W., Beekman, A. T., Smit, J. H., Zitman, F. G., Nolen, W. A., Spinhoven, P., ... Van Dyck, R. (2008). The Netherlands study of depression and anxiety (NESDA): rationale, objectives and methods. *International Journal of Methods in Psychiatric Research*, *17*, 121-140.

Plewis, I. (1996). Statistical methods for understanding cognitive growth: A review, a synthesis and an application. *British Journal of Mathematical and Statistical Psychology*, *49*, 25-42.

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

R Development Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org`

Rindskopf, D. (1984). Structural equation models: Empirical identification, heywood cases, and related problems. *Sociological Methods and Research*, *13*, 109-119.

Rizopoulos, D. (2006). ltm: An `R` package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1-25.

Robinson, P. M. (1974). Identification, estimation and large-sample theory for regressions containing unobservable variables. *International Economic Review*, *15*(3), 680-692.

SAS Institute Inc. (2011). `SAS/STAT` software, version 9.3 [Computer software manual]. Cary, NC. (http://www.sas.com)

SAS Institute Inc. (2013). *Base SAS 9.4 procedures guide.* Cary, NC: SAS Institute Inc.

Schonemann, P. H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika*, *36*(21).

Schonemann, P. H., & Wong, M. M. (1972). Some new results on factor indeterminacy. *Psychometrika*, *37*(61).

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Sherman, M., & le Cessie, S. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Communications in Statistics - Simulation and Computation*, *26*, 901-925.

Sherman, M., & Le Cessie, S. (1997). A comparison between bootstrap methods and Generalized Estimating Equations for correlated outcomes in Generalized Linear Models. *Communications in Statistics - Simulation and Computation*, *26*, 901-925.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, *35*, 137-167.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models.* Boca Raton, FL: Chapman & Hall.

Sommer, A., Katz, J., & Tarwotjo, I. (1984). Increased risk of respiratory disease and diarrhea in children with preexisting mild vitamin A deficiency. *American Society for Clinical Nutrition*, *40*, 1090-1095.

Spinhoven, P., De Rooij, M., Heiser, W., Penninx, B. W. J. H., & Smit, J. (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.

Spinhoven, P., Penelo, E., De Rooij, M., Penninx, B. W., & Ormel, J. (2013). Reciprocal effects of stable and temporary components of neuroticism and affective disorders: results of a longitudinal cohort study. *Psychological Medicine*, *44*, 337-348.

Stapleton, D. C. (1978). Analyzing political participation data with a MIMIC model. *Sociological Methodology*, *9*, 52-74.

Takane, Y. (1998). Visualization in ideal point discriminant analysis. In J. Blasius & M. J. Greenacre (Eds.), *Visualization of categorical data* (p. 441-459). New York: Academic press.

Takane, Y., Bozdogan, H., & Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, *52*, 371-392.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3), 393-408.

Ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, *67*(5), 1167-1179.

Ter Braak, C. J. F., & Verdonschot, P. F. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences*, *57*(3), 1015-1621.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (p. 73-140). Mahwah, NJ: Erlbaum.

Thurstone, L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, *1*, 31-65.

Van der Heijden, P. G. M., Mooijaart, A., & Takane, Y. (1994). Correspondence analysis and contingency models. In M. J. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences* (p. 79-111). New York: Academic Press.

Vermunt, J. K., Rodrigo, M. F., & Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods and Research*, *30*, 170-196.

Von Oertzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology*, *63*, 627-646.

Wei, L., & Stram, D. (1988). Analyzing repeated measurements with possibly missing observations by modeling marginal distributions. *Statistics in Medicine*, *7*, 139-148.

Wei, X. (2012). %PROC_R: A SAS macro that enables native R programming in the base SAS environment. *Journal of Statistical Software*, *46*.

Worku, H. M., & De Rooij, M. (2017a). Properties of ideal point classification models for bivariate binary data. *Psychometrika*, *82*(2), 308-328.

Worku, H. M., & De Rooij, M. (2017b). **mldm***: An R package for analyzing multiple binary responses using a multivariate logistic distance model.* (Software Manual)

Worku, H. M., & De Rooij, M. (2018). A multivariate logistic distance model for the analysis of multiple binary responses. *Journal of Classification*, *35*, 1-23. (`https://doi.org/10.1007/s00357-018-9251-4`)

Yee, T. W. (2010). The **VGAM** package for categorical data analysis. *Journal of*

Statistical Software, *32*(10), 1-34. Retrieved from `http://www.jstatsoft.org/v32/i10/`

Yu, H. T., & De Rooij, M. (2013). Model selection for trend vector model. *Journal of Classification*, *30*, 338-369.

Zhao, L. P., & Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, *77*, 642-648.

Ziegler, A. (2011). *Generalized estimating equations*. New York: Springer.

Ziegler, A., & Arminger, G. (1995). Analyzing the employment status with panel data from GSOEP - a comparison of the MECOSA and the GEE1 approach for marginal models. *Vierteljahreshefte zur Wirtschaftsforschung*, *64*, 72-80.

Ziegler, A., Kastner, C., & Blettner, M. (1998). The Generalized Estimating Equations: An Annotated Bibliography. *Biometrical Journal*, *40*(2), 115-139.