Missing Observations in Paired Comparison Data - supplementary material

R. Dittrich * B. Francis † R. Hatzinger *
W. Katzenbeisser *

14.05.2011

A. An algorithm to generate the C matrix

An algorithm to generate the **C** matrices may also be useful and is given here. Computationally, it is useful to generate the matrices $\mathbf{C}_{[\cdot]}$ from vectors. Let such a vector be denoted as $s_{[\cdot]}$ of length 2^{ℓ}. Each element of the vector $s_{[\cdot]}$ identifies the cells of $\mathbf{C}_{[\cdot]}$ which take the value 1, with the position of the element giving the column of $\mathbf{C}_{[\cdot]}$ and the value of the element giving the row. For instance, returning to the earlier example, for $\mathbf{C}_{[23]}$ the corresponding $s_{[23]} = (1, 1, 2, 2, 3, 3, 4, 4)$, or for $\mathbf{C}_{[13][23]}$ it is $s_{[13][23]} = (1, 1, 1, 1, 2, 2, 2, 2)$. In general, given an arbitrary number of comparisons, the entries in $s_{[\cdot]}$ can easily be computed using the following steps: (i) For each missing block convert the matrix $\mathbf{Y}$ with rows $y = (y_{obs}, y_{mis})$ (see Table 1) to have binary entries such that the response 1 is coded as 0 and -1 is coded as 1. (ii) Delete the columns with missing entries (the starred entries in Table 1). (iii) Consider each row of the resulting matrix as a binary number, convert it to a decimal number and add 1. This gives the vectors $s_{[\cdot]}$ distinct for every missing block.

Using the $s_{[\cdot]}$ described above, the necessary summations of the probabilities for each block separately can be applied immediately:

$$P(y_{[\cdot]k}) = \sum_{i:s_{[\cdot]i}=k} \gamma_i\,, \qquad i = 1, \ldots, 2^{\ell}\,, \tag{0.1}$$

*Address for correspondence: Regina Dittrich, WU Vienna University of Economics and Business, Institute for Statistics and Mathematics, Augasse 2-6, 1090 Vienna, Austria, Regina.Dittrich@wu.ac.at

†Lancaster University, UK

where $k = 1, \dots, 2^{\ell-\kappa}$. In R, the computation can be performed using the R function `tapply(gamma, s, sum)` where the elements of the vector `gamma` are summed according to the values of `s`.

	observed data				complete data $(\mathbf{Y} \circ \mathbf{R})$					
missing block	*number of cases*	*observed pattern*			$y = (y_{obs}, y_{mis})$			*missing indicators*		
		(12)	(13)	(23)	(12)	(13)	(23)	(12)	(13)	(23)
[23]	$n_{(1,1,na)}$	1	1	*na*	1	1	$1*$	0	0	1
					1	1	$-1*$	0	0	1
	$n_{(1,-1,na)}$	1	-1	*na*	1	-1	$1*$	0	0	1
					1	-1	$-1*$	0	0	1
	$n_{(-1,1,na)}$	-1	1	*na*	-1	1	$1*$	0	0	1
					-1	1	$-1*$	0	0	1
	$n_{(-1,-1,na)}$	-1	-1	*na*	-1	-1	$1*$	0	0	1
					-1	-1	$-1*$	0	0	1

Table 1: Data structure for one missing block: Example for Y_{23} missing (extract from Table 1 in main paper)

B. Simulations

Three MNAR models were specified for the simulations, model A with three items and models B and C with four items. For models A and B we defined the first item to produce nonignorable missing values using $\alpha_1 = -0.5$ and $\beta_1 = 0.2$; all other α and β values were therefore set to zero. For model C additionally item 2 was specified to produce nonignorable missing values using $\alpha_2 = -0.7$ and $\beta_2 = -0.2$. The λs for model A were fixed at $(-0.3, 0.3, 0)$ and for models B and C at $(-0.5, 0.3, 0.2, 0)$, respectively.

Using these parameter values, we produced "true" (expected) probabilities for each pattern. Based on these probabilities, random samples were drawn from a multinomial distribution to obtain simulated counts for the paired comparison patterns. We used sample sizes of 100, 500, 1000, and 2000. Each combination of models A, B, and C with the four different sample sizes was replicated 10000 times. For each simulation we assessed sensitivity by fitting two models - an "incorrect" MCAR model (without βs) and a "correct" MNAR model (including the βs). The corresponding MSEs for each fitted model are reported in Tables 2, 3, and 4. The mean square error (MSE) of the estimated parameters is defined as the mean squared difference between the estimates of the parameters and the true parameter values. We also report the maximum number of iterations, which is the largest number of iterations used by the fitting procedure over all 10000 simulations.

		N = 100		N = 500		N = 1000		N = 2000	
parameters		MSE		MSE		MSE		MSE	
true values		MNAR	MCAR	MNAR	MCAR	MNAR	MCAR	MNAR	MCAR
λ_1	-0.3	.0129	.0120	.0025	.0029	.0012	.0018	.0006	.0012
λ_2	0.3	.0091	.0087	.0018	.0020	.0009	.0011	.0004	.0007
α_1	-0.5	.1272	.0294	.0058	.0109	.0028	.0087	.0014	.0074
β_1	0.2	.1687		.0136		.0065		.0032	
max. iter.		44	11	16	11	15	10	14	9

Table 2: Model A: 3 Items, Item 1 missing

		N = 100		N = 500		N = 1000		N = 2000	
parameters		MSE		MSE		MSE		MSE	
true values		MNAR	MCAR	MNAR	MCAR	MNAR	MCAR	MNAR	MCAR
λ_1	$-.5$	.0105	.0104	.0021	.0033	.0010	.0024	.0005	.0020
λ_2	.3	.0066	.0066	.0013	.0017	.0007	.0010	.0003	.0007
λ_3	.2	.0059	.0058	.0012	.0012	.0006	.0006	.0003	.0003
α_1	$-.5$	.0309	.0247	.0037	.0130	.0018	.0113	.0009	.0107
β_1	.2	.0407		.0053		.0027		.0013	
max. iter.		32	12	14	12	15	12	15	11

Table 3: Model B: 4 Items, Item 1 missing

		N = 100		N = 500		N = 1000		N = 2000	
parameters		MSE		MSE		MSE		MSE	
true values		MNAR	MCAR	MNAR	MCAR	MNAR	MCAR	MNAR	MCAR
λ_1	$-.5$	.0131	.0101	.0025	.0019	.0013	.0010	.0007	.0006
λ_2	.3	.0082	.0090	.0016	.0029	.0009	.0022	.0005	.0018
λ_3	.2	.0071	.0072	.0014	.0018	.0007	.0012	.0003	.0008
α_1	$-.5$	.0542	.0265	.0455	.0141	.0448	.0126	.0441	.0117
α_2	$-.7$	.4199	.0229	.0127	.0104	.0082	.0088	.0062	.0080
β_1	.2	.0316		.0061		.0033		.0020	
β_2	$-.2$	.4038		.0071		.0033		.0017	
max. iter.		73	14	20	14	19	13	19	13

Table 4: Model C: 4 Items, Items 1 and 2 missing

In examining the results of the simulations, we can observe three interesting features. Firstly, we see that the MSE values for $N = 100$ are considerably larger than those values for larger sample sizes, especially for alpha and beta. Secondly, for sample sizes of 500 or more, the algorithm shows uniform behaviour in MSE, with very small MSEs being reported for all parameters and MSE declining in size as sample size increases. It is also interesting to note that the estimates for the misspecified MCAR model broadly agree with the true parameter values from the simulated MNAR model for those parameters that are estimated. The parameter estimates of the λs are relatively unaffected by misspecification of the model. Finally, in terms of stability, we notice that the maximum number of iterations declines as the sample size increases; in general between 15 and 20 iterations of the maximisation algorithm are needed.

C. Parameter estimates and standard errors for the four models with and without dependencies

	CC-outcome model				MCAR model 2			
	without dependencies		with dependencies		without dependencies		with dependencies	
effect	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.
λ_{ST}	.262	.045	.278	.054	.271	.042	.288	.051
λ_{CM}	−.084	.045	−.106	.052	−.073	.042	−.081	.049
λ_{AC}	.152	.045	.142	.047	.164	.042	.154	.045
λ_{SU}	.105	.045	.115	.051	.109	.042	.122	.048
λ_{VA}	0	—	0	—	0	—	0	—
$\theta_{12.13}$			−.116	.078			−.129	.074
$\theta_{12.14}$			.185	.084			.197	.078
$\theta_{12.15}$			−.072	.080			−.025	.075
$\theta_{13.14}$			−.068	.076			−.077	.072
$\theta_{13.15}$			−.167	.073			−.163	.069
$\theta_{14.15}$			.019	.076			−.013	.072
$\theta_{12.23}$			.261	.086			.232	.080
$\theta_{12.24}$			−.444	.087			−.411	.081
$\theta_{12.25}$			.107	.084			.092	.080
$\theta_{23.24}$			.064	.085			.037	.079
$\theta_{23.25}$			.160	.076			.174	.072
$\theta_{24.25}$			−.208	.083			−.175	.078
$\theta_{13.23}$			.036	.074			.065	.070
$\theta_{13.34}$			−.209	.073			−.221	.069
$\theta_{13.35}$			−.092	.074			−.107	.071
$\theta_{23.34}$			−.002	.073			.021	.069
$\theta_{23.35}$			.178	.074			.185	.071
$\theta_{34.35}$			.115	.074			.126	.071
$\theta_{14.24}$			−.013	.083			−.045	.077
$\theta_{14.34}$			.140	.075			.115	.071
$\theta_{14.45}$			−.121	.073			−.119	.070
$\theta_{24.34}$			.028	.074			.018	.070
$\theta_{24.45}$			.125	.076			.105	.072
$\theta_{34.45}$			.145	.073			.145	.069
$\theta_{15.25}$			.078	.072			.030	.069
$\theta_{15.35}$			.050	.073			.034	.070
$\theta_{15.45}$			.045	.072			.045	.068
$\theta_{25.35}$			−.096	.073			−.106	.070
$\theta_{25.45}$			.017	.072			.028	.069
$\theta_{35.45}$			−.046	.072			−.033	.069
α					−3.642	.130	−3.642	.130

Table 5: Parameter estimates and standard errors for complete case (CC-outcome) models and MCAR models with a common α, estimated with and without dependence parameters.

	MCAR model 3				MNAR model			
	without dependencies		with dependencies		without dependencies		with dependencies	
effect	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.
λ_{ST}	.271	.042	.288	.051	.268	.043	.273	.052
λ_{CM}	−.073	.042	−.081	.049	−.074	.043	−.087	.050
λ_{AC}	.164	.042	.154	.045	.163	.042	.147	.045
λ_{SU}	.109	.042	.122	.048	.115	.042	.129	.049
λ_{VA}	0	—	0	—	0	—	0	—
$\theta_{12.13}$			−.129	.074			−.116	.075
$\theta_{12.14}$			.197	.078			.200	.078
$\theta_{12.15}$			−.025	.075			−.028	.075
$\theta_{13.14}$			−.077	.072			−.073	.072
$\theta_{13.15}$			−.163	.069			−.155	.069
$\theta_{14.15}$			−.013	.072			−.007	.072
$\theta_{12.23}$			.232	.080			.231	.079
$\theta_{12.24}$			−.411	.081			−.401	.081
$\theta_{12.25}$			.092	.080			.104	.079
$\theta_{23.24}$			.037	.079			.040	.078
$\theta_{23.25}$			.174	.072			.171	.072
$\theta_{24.25}$			−.175	.078			−.173	.078
$\theta_{13.23}$			.065	.070			.056	.070
$\theta_{13.34}$			−.221	.069			−.221	.069
$\theta_{13.35}$			−.107	.071			−.103	.071
$\theta_{23.34}$			.021	.069			.020	.069
$\theta_{23.35}$			.185	.071			.179	.071
$\theta_{34.35}$			.126	.071			.128	.071
$\theta_{14.24}$			−.045	.077			−.041	.078
$\theta_{14.34}$			.115	.071			.111	.071
$\theta_{14.45}$			−.119	.070			−.117	.070
$\theta_{24.34}$			.018	.070			.017	.070
$\theta_{24.45}$			.105	.072			.109	.072
$\theta_{34.45}$			.145	.069			.145	.069
$\theta_{15.25}$			.030	.069			.029	.069
$\theta_{15.35}$			.034	.070			.037	.070
$\theta_{15.45}$			.045	.068			.045	.068
$\theta_{25.35}$			−.106	.070			−.105	.070
$\theta_{25.45}$			.028	.069			.027	.069
$\theta_{35.45}$			−.033	.069			−.033	.069
α_1	−2.087	.240	−2.087	.240	−2.410	.380	−2.427	.383
α_2	−1.795	.220	−1.795	.220	−1.850	.232	−1.856	.233
α_3	−1.851	.223	−1.851	.223	−1.848	.228	−1.866	.228
α_4	−1.908	.227	−1.908	.227	−2.212	.323	−2.189	.323
α_5	−1.525	.207	−1.525	.207	−1.543	.218	−1.481	.217
β_1					.570	.397	.552	.391
β_2					−.416	.221	−.305	.219
β_3					.395	.234	.276	.233
β_4					.656	.329	.566	.332
β_5					−.084	.240	.120	.251

Table 6: Parameter estimates and standard errors for MCAR models with different αs for all items and MNAR models with different αs and βs for all items, estimated with and without dependence parameters.