

D. Additional Simulation Study to Chapter 3

We provide an additional simulation study under settings similar to Study I in Section 3.6.1, but with smaller values of J to mimic educational tests with relatively smaller item sizes.

D.1. Settings

We consider two settings, with (1) $N = 2,000$, $J = 50$, and (2) $N = 4,000$, $J = 100$. The rest of the simulation setting is exactly the same as that of Study I. These two settings are referred to as settings C.S1 and C.S2, respectively. For each setting, 100 independent data sets are simulated.

D.2. Results

The analysis is conducted using the parallel tempering MCMC algorithm described above. For each data set, we run 10,000 iterations, with the first 3,000 iterations as the burn-in. The results are based on the posterior samples from the last 7,000 iterations.

Our results are given in Tables *Appendix D.2.(a)* through *Appendix D.2.(d)*. The results are similar to those from Study I. Table *Appendix D.2.(a)* gives the AUC values for the classification of test takers and items. Tables *Appendix D.2.(b)* and *Appendix D.2.(c)* evaluate the performance of the local FDR control and local FNR control procedures for the classification of test takers and items, respectively. More specifically, Table *Appendix D.2.(b)* gives the corresponding FDP values when controlling the local FDR at 1%, 5%, and 10% levels for the test takers. Table *Appendix D.2.(c)* gives the FNP values when controlling the local FNR at 1%, 5%, and 10% levels for the items. Finally, the bias and variance for the posterior mean estimator are given in Table *Appendix D.2.(d)*.

AUC	Test taker				Item			
	C.S1		C.S2		C.S1		C.S2	
	Reduced	Full	Reduced	Full	Reduced	Full	Reduced	Full
25%	0.934	0.936	0.956	0.962	0.937	0.944	0.954	0.961
50%	0.955	0.954	0.960	0.971	0.964	0.963	0.972	0.976
75%	0.969	0.967	0.971	0.978	0.973	0.975	0.978	0.981

Appendix D.2.(a) Overall classification performance based on the posterior means of ξ_i and η_j . For each model, each setting, and each target (test taker/item), we show the 25%, 50%, and 75% quantiles of the AUCs of the corresponding ROC curves from 100 independent data sets.

FDP	C.S1						C.S2					
	Reduced			Full			Reduced			Full		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
25%	0.008	0.032	0.089	0.008	0.035	0.085	0.007	0.028	0.072	0.008	0.029	0.069
50%	0.012	0.047	0.092	0.011	0.046	0.091	0.011	0.043	0.078	0.009	0.037	0.073
75%	0.013	0.051	0.098	0.013	0.053	0.099	0.013	0.047	0.084	0.012	0.044	0.087

Appendix D.2.(b) Local FDR control for test takers. For each model, each setting, and each local FDR target (1%/5%/10%), we show the 25%, 50%, and 75% quantiles of the FDPs of the corresponding classifications from 100 independent data sets.

FNP	S1						S2					
	Reduced			Full			Reduced			Full		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
25%	0.009	0.033	0.063	0.007	0.032	0.061	0.007	0.029	0.059	0.006	0.031	0.062
50%	0.012	0.037	0.068	0.009	0.036	0.067	0.009	0.038	0.067	0.007	0.036	0.065
75%	0.013	0.046	0.071	0.010	0.043	0.069	0.012	0.045	0.072	0.012	0.041	0.071

Appendix D.2.(c) Local FNR control for items. For each model, each setting, and each local FNR target (1%/5%/10%), we show the 25%, 50%, and 75% quantiles of the FNPs of the corresponding classifications from 100 independent data sets.

C.S1	Reduced model						C.S2	Reduced model					
	π_1	π_2	σ_{11}	μ_1	ω_{11}	δ		π_1	π_2	σ_{11}	μ_1	ω_{11}	δ
Bias	0.11	-0.02	-0.05	-0.11	0.17	0.22	Bias	0.05	0.15	-0.04	-0.13	0.20	-0.07
Variance	0.16	0.09	0.32	0.43	0.17	0.30	Variance	0.21	0.11	0.39	0.25	0.22	0.22
C.S1	Full model						C.S2	Full model					
	π_1	π_2	σ_{11}	μ_1	ω_{11}	δ		σ_{22}	σ_{12}	μ_2	ω_{22}	ω_{12}	κ
Bias	-0.11	-0.08	0.07	0.24	-0.08	0.11		-0.12	-0.04	-0.12	0.14	-0.08	-0.16
Variance	0.17	0.19	0.32	0.32	0.21	0.30		0.11	0.17	0.15	0.23	0.00	0.37
C.S2	Full model						C.S2	Full model					
	π_1	π_2	σ_{11}	μ_1	ω_{11}	δ		σ_{22}	σ_{12}	μ_2	ω_{22}	ω_{12}	κ
Bias	-0.04	0.06	0.08	-0.15	-0.17	-0.15		-0.04	-0.11	-0.15	0.11	-0.04	-0.11
Variance	0.21	0.18	0.23	0.29	0.31	0.35		0.07	0.19	0.12	0.12	0.18	0.46

Appendix D.2.(d) Accuracy of the posterior mean estimator of the global parameters. The bias and variance for the posterior mean estimator are calculated based on the 100 replications.