Online Appendix C

In the main paper, we present evidence that, when 4PL data are incorrectly fit to the 3PL, trait estimation bias is highly sensitive to the choice of Bayesian priors used in item parameter estimation. In this appendix, we demonstrate that these results are not unique to the `mirt` software. Using the same data-generating procedures described in the main paper, we estimated 1PL, 2PL, and 3PL item parameters using Hamiltonian Monte Carlo (HMC) with the no-U-turn sampling algorithm (NUTS; Hoffman & Gelman, 2014) as implemented in the `rstan` (Stan Development Team, 2016) package for `R`. We used the same priors on all parameters as those used in the original item parameter estimation. Namely, $\theta \sim N(0, 1)$ for all models and $\text{logit}(c) \sim N(-1.8, 1)$ for the 3PL. There are two major differences in the MML (`mirt`) and HMC (`rstan`) estimation methods. First, whereas MML estimates item parameters by first integrating over a distribution of person parameters, HMC produces estimates of both the item and person parameters (although we only used the estimated item parameters in these simulations). Second, whereas MML optimizes the marginal likelihood function directly, HMC repeatedly samples from the joint posterior distribution of the parameter space. In HMC, the mean sampled value for each parameter is taken to be the parameter estimate. Despite these differences, we computed trait estimates in the same fashion as the MML simulations described in the main paper. Namely, we treated the HMC item parameter estimates as known and computed EAP2 trait estimates for 26,000 response vectors generated from the 4PL. In these follow-up simulations, all item parameters were estimated based on exactly one of the $N = 20,000$ data sets that were used to compute MML estimates of the 1PL, 2PL, and 3PL. Due to the increased time needed to run models in `rstan`, only one set of item parameters was estimated for each model. All IRT parameters were estimated using four chains of length 1000, wherein the first 500 draws of each chain were discarded. The item parameters obtained with `rstan` were used to compute EAP2 trait estimates on the same set of 26,000 response vectors described in the main paper. Notably, the trait estimation bias associated with each set of item parameter estimates was highly similar to that obtained by the original `mirt`-estimated parameters. The mean absolute difference between the 26,000 EAP2 estimates computed from the MML and HMC item parameter estimates equalled .004, .001, and .006 for the 1PL, 2PL, and 3PL. Because the MML and HMC estimation methods led to very similar bias results, we can conclude that the bias values observed for traits computed from these misspecified models are not unique to the `mirt` software. These results indicate that the pattern of bias observed for the 3PL is a consequence of the 3PL model itself and may be manipulated by the choice of Bayesian priors used in 3PL item parameter estimation.

References

Hoffman, M. D., & Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Stan Development Team. 2016. *RStan: the R interface to Stan.* R package version 2.14.1.