Research Report

Bias Correction for the Maximum Likelihood Estimate of Ability

Jinming Zhang

Research & Development

September 2005
RR-05-15

Bias Correction for the Maximum Likelihood Estimate of Ability

Jinming Zhang
ETS, Princeton, NJ

September 2005

Abstract

Lord's bias function and the weighted likelihood estimation method are effective in reducing the bias of the maximum likelihood estimate of an examinee's ability under the assumption that the true item parameters are known. This paper presents simulation studies to determine the effectiveness of these two methods in reducing the bias when the item parameters are unknown. The simulation results show that Lord's bias function and the weighted likelihood estimation method might not be as effective in bias reduction in the 3PL cases when item parameters are unknown as they are when the true item parameters are given. Algorithms and methods for obtaining the global maximum value of a likelihood function or a weighted likelihood function are discussed in this paper.

Key words: Item response theory (IRT), MLE, WLE, bias reduction, bisection method

1. Introduction

Item response theory (IRT) describes the interaction between examinees and test items using probabilistic models, typically under an assumption that the performance of examinees on a test can be explained by one latent ability. When IRT models are used to estimate the abilities of examinees in a target sample, the item parameters in these models are usually estimated first. Then, treating these estimates as fixed, ability parameters are estimated and further statistical inferences are made. A calibration sample, which is used to estimate item parameters, may be the same as the target sample itself, in which examinees' abilities are to be estimated. This paper mainly focuses on the case where a calibration sample and a target sample are different from each other. For example, in computer-based testing (CBT) programs, including computerized adaptive testing (CAT) and linear form (on-the-fly) testing, test items are typically calibrated first using response data from a pretest or a field test. These items are put into an item pool and the estimated item parameters are treated as fixed to estimate the abilities of examinees in the later tests. However, item parameter estimates are contaminated by statistical errors. If these errors are severe but ignored, the statistical inferences based on the estimated abilities could be misleading. Therefore, it is very important to investigate how large the impact of statistical errors of item parameter estimates is on the estimation of examinees' abilities.

The maximum likelihood estimation method is the most popular method in the estimation of unknown parameters in a statistical model. As pointed out by Lord (1983, 1986), even assuming true item parameters are known, the maximum likelihood estimate (MLE) of an examinee's ability still has bias. Given true item parameters, Lord used Taylor's expansion of the likelihood equation and obtained an approximation for the bias and its standard error formulae for the MLE of ability in the context of the three-parameter logistic (3PL) and two-parameter logistic (2PL) models. Samejima (1993a, 1993b) further developed these formulae for more general response models, where item responses are discrete. Warm (1989) used the weighted likelihood estimation method and Lord's bias function to estimate ability parameters and proved that the weighted likelihood estimate (WLE) is less biased than the MLE with the same asymptotic variance and normal distribution. Assuming the true item parameters are known, these methods are effective in reducing bias.

In reality, both item and ability parameters are unknown. As mentioned earlier, it is a common practice to estimate item parameters first and then treat the estimates as if they were the true quantities in estimating ability parameters. That is, the MLE of an ability parameter is obtained by assuming estimated item parameters are fixed as substitutes for true parameters. This paper examines the effectiveness of Lord's bias function and the weighted likelihood estimation method in bias reduction when item parameters are estimated. The most widely used calibration programs at present are BILOG (Mislevy & Bock, 1982) and PARSCALE (Muraki & Bock, 1991, 1997). Both programs use the marginal maximum likelihood estimation approach and an EM algorithm (Bock & Aitkin, 1981) to estimate item parameters. In this paper, the NAEP BILOG/PARSCALE program, an item parameter estimation program that combines BILOG and PARSCALE (see Allen, Carlson, & Zelenak, 1999), was used to estimate item parameters. For convenience, PARSCALE was used to represent the NAEP BILOG/PARSCALE program in this paper. Simulation studies were conducted to compare MLE with WLE and MLE with Lord's bias correction (denoted as MLE-LBC) with respect to the bias and the root mean squared errors (RMSE) for estimated abilities. To check the impact of estimated item parameters as substitutes for true parameters on the methods of bias reduction for ability estimates, two scenarios were considered in the simulation studies: True item parameters are (a) given; or (b) need to be estimated.

2. MLE, WLE, and Lord's Bias Function

Suppose a test consists of n dichotomous items. Let $\mathbf{y} = (y_1, \ldots, y_n)$ be the response vector of an examinee with $y_i = 1$ (correct) or $y_i = 0$ (incorrect), $i = 1, \ldots, n$. The item response function (IRF) of a 3PL model is

$$P_i(\theta) = P(y_i = 1|\theta) = c_i + (1 - c_i)\frac{1}{1 + \exp\{-1.7a_i(\theta - b_i)\}}, \qquad (1)$$

where a_i, b_i, and c_i are the item discrimination, difficulty, and guessing parameters, respectively. When the guessing parameter is fixed at zero, the 3PL model becomes a 2PL model.

The MLE of examinee's ability is commonly used in practice (Birnbaum, 1968; Wang & Vispoel, 1998; Yi, Wang, & Ban, 2001). Under the assumptions of *local* or *conditional* independence (Lord, 1980), the likelihood function for the response vector $\mathbf{y}$ is

$$L(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} P_i^{y_i}(\theta)\, Q_i^{1-y_i}(\theta), \tag{2}$$

where $Q_i(\theta) = 1 - P_i(\theta)$. If item parameters (a_i, b_i, c_i) in these models are known, the MLE $\hat{\theta}$ for the ability parameter is defined as the value of θ that maximizes (2). In practice, $\hat{\theta}$ is often found at the zero of the likelihood equation; that is, $\hat{\theta}$ satisfies

$$\sum_{i=1}^{n} \left(\frac{y_i - P_i(\theta)}{P_i(\theta) Q_i(\theta)} \right) P_i'(\theta) = 0, \tag{3}$$

where $P_i'(\theta)$ is the first derivative of $P_i(\theta)$ with respect to θ (see Lord, 1980). Let

$$I(\theta) = \sum_{i=1}^{n} \frac{(P_i'(\theta))^2}{P_i(\theta) Q_i(\theta)}$$

be the Fisher test information function. The variance of the MLE of θ is $\mathrm{Var}(\hat{\theta}) = I(\hat{\theta})^{-1}$. Note that the likelihood equation may have no root or several roots. A root of a likelihood equation is also called a *critical point.* Figure 1 shows the curves of a log likelihood function and its likelihood equation function from our simulation studies (see the next section for details about the simulated test). The likelihood equation has three roots (critical points) in this example. Clearly, a critical point is not necessarily the point achieving the global maxima and, thus, is not necessarily the MLE of θ.

Given item parameters, Lord (1983) obtained the following bias function for the MLE of θ,

$$B_1(\theta) = \frac{1.7}{I^2(\theta)} \sum_{i=1}^{n} a_i I_i(\theta) \left(P_i^*(\theta) - \frac{1}{2} \right), \tag{4}$$

where

$$P_i^*(\theta) = \frac{P_i(\theta) - c_i}{1 - c_i} \equiv \frac{1}{1 + \exp\{-1.7 a_i(\theta - b_i)\}} \quad \text{and} \quad I_i(\theta) = \frac{(P_i'(\theta))^2}{P_i(\theta) Q_i(\theta)}.$$

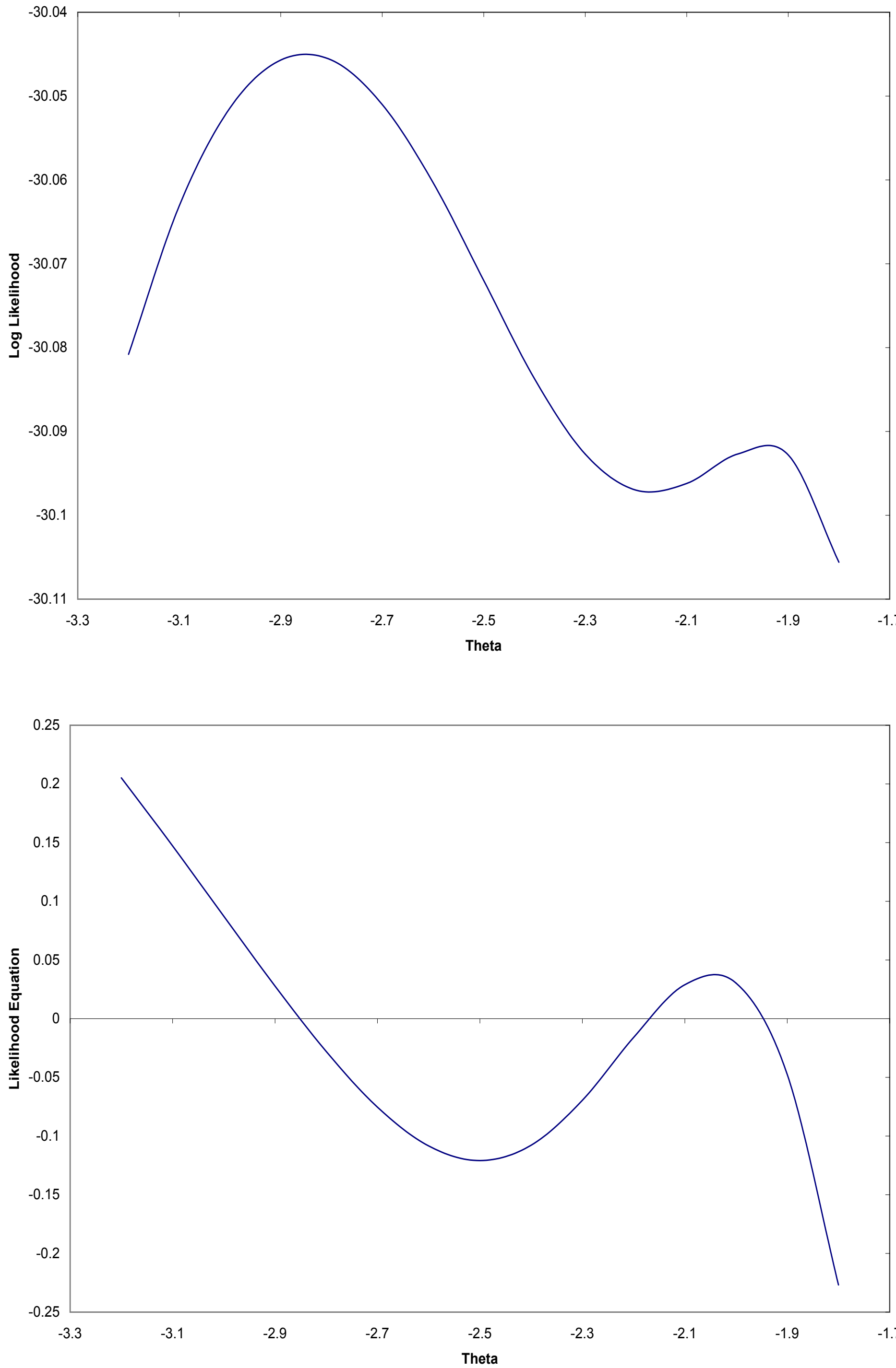

Figure 1. **A special log likelihood function curve with true ability being -2.6 for 50 3PL items with estimated item parameters and its likelihood equation with three roots.**

The MLE-LBC of θ is defined as

$$\hat{\theta}_c = \hat{\theta} - B_1(\hat{\theta}).$$

The bias of $\hat{\theta}_c$, BIAS($\hat{\theta}_c$), is $o(n^{-1})$—that is, $\lim_{n\to\infty} n\text{BIAS}(\hat{\theta}) = 0$—while BIAS($\hat{\theta}$) is $O(n^{-1})$—that is, nBIAS($\hat{\theta}$) are bounded for all n—under the assumption that the true values of item parameters are known.

Based on Lord's work, Warm (1989) proposed the weighted likelihood estimation method and proved that the weighted likelihood estimate (WLE) of ability is less biased than the MLE with the same asymptotic variance and normal distribution. The WLE, $\hat{\theta_w}$, is defined as the value of θ that maximizes

$$f(\theta)L(\mathbf{y} \mid \theta) = f(\theta) \prod_{i=1}^{n} P_i^{y_i}(\theta)\, Q_i^{1-y_i}(\theta), \tag{5}$$

where $f(\theta)$ is a suitably chosen function satisfying

$$\frac{\partial \ln f(\theta)}{\partial \theta} = -B_1(\theta)I(\theta).$$

Or

$$f(\theta) = \exp\left\{ -\int_{-\infty}^{\theta} B_1(x)I(x)dx \right\}. \tag{6}$$

Therefore, $\hat{\theta_w}$ satisfies the following weighted likelihood equation,

$$\sum_{i=1}^{n} \left(\frac{y_i - P_i(\theta)}{P_i(\theta)Q_i(\theta)} \right) P_i'(\theta) - B_1(\theta)I(\theta) = 0. \tag{7}$$

The ability parameter is iteratively estimated by the bisection method with 15 iterations (Warm, 1989). However, the weighted likelihood equation can have several roots. Thus, a root of the weighted likelihood equation (7) is not necessarily the WLE of θ, just as a root of the likelihood equation (3) is not necessarily the solution of MLE. Since the weight function $f(\theta)$ in (6) has no explicit closed-form, numerical integration is needed in order to check whether or not a root from the weighted likelihood equation actually maximizes the weighted likelihood function (5), which is more difficult to check than checking if the root from the likelihood equation (3) maximizes the likelihood function (2). Warm

(1989, p. 431) claimed that only $\partial \ln f(\theta)/\partial\theta$ is necessary in order to solve for the WLE. Unfortunately, it is only true when searching for a root of the weighted likelihood equation, but it is not true when trying to find the WLE of θ.

The ability of a randomly selected examinee is a latent variable, and thus is scale-free in theory. In practice, it is common to assume that the distribution of abilities for a fixed examinee population is standard normal. The item difficulty parameters typically range from -2.5 to 2.5. It is almost impossible for such a test to distinguish two very high or two very low ability examinees, say with abilities of -4 and -5. In fact, it is not important to do so since few examinees actually achieve these values. Thus this paper only considers ability scores on a closed interval, $[-d, d]$, where d is a fixed integer (e.g., $d = 4$). That is, this paper only searches for the absolute or global maximum value on a closed interval. Note that Lord (1983) also assumed that all item and ability parameters were bounded. Because the roots of the likelihood equation (3) and the weighted likelihood equation (7) may not be unique, it is important to find all critical points. To obtain the global maxima on a closed interval, it is necessary to compare all the likelihood function values or the weighted likelihood function values at critical points and at the endpoints of the closed interval. Note that the MLE exists even for the perfect or zero scores on a closed interval (the maximum value is achieved at one of the endpoints).

Newton-Raphson's method can be used to search for the roots (critical points) of the likelihood or weighted likelihood equations. However, this method fails to provide a sequence of approximations that converge to a solution in many cases in simulation studies. Therefore, the method of bisection is used in the Fortran program developed in this paper. The comparison of this program with PARSCALE, which uses the Newton-Raphson method in both its MLE and WLE (Muraki & Bock, 1997), is under way. Further studies will adopt a hybrid algorithm that utilizes a combination of bisection and Newton-Raphson methods or will directly use an optimization method, such as the golden section method, to search for the value where the likelihood function or the weighted likelihood function takes on the maximum value. All of these studies will be reported on in a later paper.

3. Simulation Studies

Simulation studies were conducted to compare MLE, WLE, and MLE-LBC. In the simulation, the true item discrimination parameters were generated independently from a normal distribution with mean 1.0, standard deviation 0.8, upper bound 1.8, and lower bound 0.4. The true item difficulty parameters were generated independently from a normal distribution with mean 0.0, standard deviation 1.0, upper bound 2.5, and lower bound -2.5. The guessing parameters in the 3PL model that is being used were generated independently from the $(0.0, 0.4)$ uniform distribution. As mentioned in the last section, $\text{BIAS}(\hat{\theta}_c)$ is $o(n^{-1})$, where n is the number of items. The number of items considered by Lord (1983) is 90. In this simulation, the number of items is set to be 50, 100, or 200. All items in a given test in the simulation are modeled either by 3PL or 2PL models. In this way, the effect of different models on the estimation of ability can be examined. In practice, a multiple-choice item is usually modeled by a 3PL model and a dichotomously-scored constructed-response item by a 2PL model.

The ability scores were chosen to be 31 evenly distributed points on $[-3, 3]$ with 1,000 students (independent replications) at each ability level. That is, the true ability scores take values from 31 concrete levels, -3.0, -2.8, $\ldots$, 2.8, 3.0, and there are 1,000 examinees (simulees) at each level. Using these ability scores and item parameters, simulated response data are then generated according to the following commonly used IRT method: Given ability score θ, first calculate the probability of answering item i correctly by the examinee with θ, $p_i = P_i(\theta)$, using the true item parameters. Then independently generate a random number r from the $(0, 1)$ uniform distribution. If $r < p_i$, then a correct response is obtained for the examinee on item i; otherwise, an incorrect response is obtained. This simulated response data set is the target data set in this paper.

As mentioned in Section 1, two situations are considered here: Item parameters are either known or not. In cases where item parameters were unknown, a separate calibration sample was generated with the same item parameters as used in generating the corresponding target response data, and with 500 or 1,000 examinees in the calibration sample whose ability scores were independently generated from the standard normal distribution. This simulated calibration sample can be regarded as a pretest or a field test sample used in practice to establish a test item pool. PARSCALE was then used to estimate item parameters with the calibration samples. The estimated item parameters were treated as fixed and used to estimate abilities for the target data. The reader may consider this simulation process to parallel the workings of linear form CBT in which the calibration and target samples are different.

In summary, these simulation studies considered four factors: (a) 31 levels of true ability, (b) three levels of numbers of items (50, 100, or 200), (c) two IRT models (2PL and 3PL), and (d) the availability of true item parameters. In the case with unknown item parameters, the calibration sample size has two levels, 500 or 1,000 simulated examinees.

The criteria used for comparisons were conditional average bias and the RMSE at each ability level. RMSE is the square root of the average of the squared deviations of estimated parameters from the true one.

Tables A1 to A6 in the appendix present the conditional average bias and the conditional RMSE of ability at each of the 31 ability levels considered in this paper when the true item parameters are known. These six tables illustrate the six cases corresponding to three levels of numbers of items across two IRT models. Figures 2 and 3 are the corresponding plots for bias and RMSE, respectively. Each figure consists of six plots corresponding to the six cases. Note that the vertical scale for the first plots is different from the rest of the five plots for an appropriate presentation. Figure 2 shows that WLE outperformed the other two methods overall in terms of average bias, whereas MLE-LBC outperformed in most cases by having the smallest RMSE, except for the cases of 50 3PL items, as seen in Figure 3. As expected, the results for 2PL models are better than the corresponding results for 3PL models with the same number of items. For details, see Tables A1-A6.

Figure 4 presents the curves representing Lord's bias function in (4) for the cases of 50 and 200 3PL items based on the true item parameters. When $|\theta| > 2$, Lord's bias function for the 50 items increases much faster than that for the 200 items. Generally, when the number of items is not large enough and the considered ability level is out of the range of the difficulty parameters, the Lord's bias function values are large, usually more than one. Consequently, Lord's bias function may overcorrect for the bias of MLE in these cases. Nevertheless, WLE and MLE-LBC, especially WLE, effectively reduced the bias of MLE in the simulation studies when the true item parameters were given.

When the item parameters were unknown, a simulated calibration sample with 500 or 1,000 examinees was used to estimate item parameters by PARSCALE. Table 1 presents sample correlations between estimated and true a, b, and c parameters separately for each kind of parameters and the average RMSE of estimated IRFs. The RMSE of an estimated IRF is defined as

$$\sqrt{\int [\hat{P}_i(\theta) - P_i(\theta)]^2 \varphi(\theta) d\theta},$$

where $P_i(\theta)$ is given by (1), $\hat{P}_i(\theta)$ is the estimate of $P_i(\theta)$ (i.e., $P_i(\theta)$ with estimated item parameters in place of true item parameters) and $\varphi(\theta)$ is the standard normal distribution. Table 2 presents the average bias and the RMSE of each kind of item parameters. Tables 1 and 2 show that the estimation results for all the cases considered in the simulation are reasonably accurate in the sense that all the correlations are fairly high, and the average RMSE of estimated IRFs and the RMSE of each kind of item parameters are relatively small. The average bias is also reasonably small except for the cases of 100 or 200 3PL items.

Tables A7 to A12 in the appendix present the conditional average bias and the conditional RMSE of ability at each of the 31 levels when the item parameters are estimated by PARSCALE using a calibration sample with 500 examinees, and Tables A13-A18 provide the results for the case with 1,000 examinees in the calibration sample. Figures 5-8 are the corresponding plots for bias and RMSE.

Lord (1983) showed that the MLE has negative bias for low ability levels and positive bias for high ability levels (i.e., outward-bias), and the magnitude of the bias is greater at low ability levels than it is at high ability levels, given true item parameters. The simulation confirmed this finding (see Figure 2).

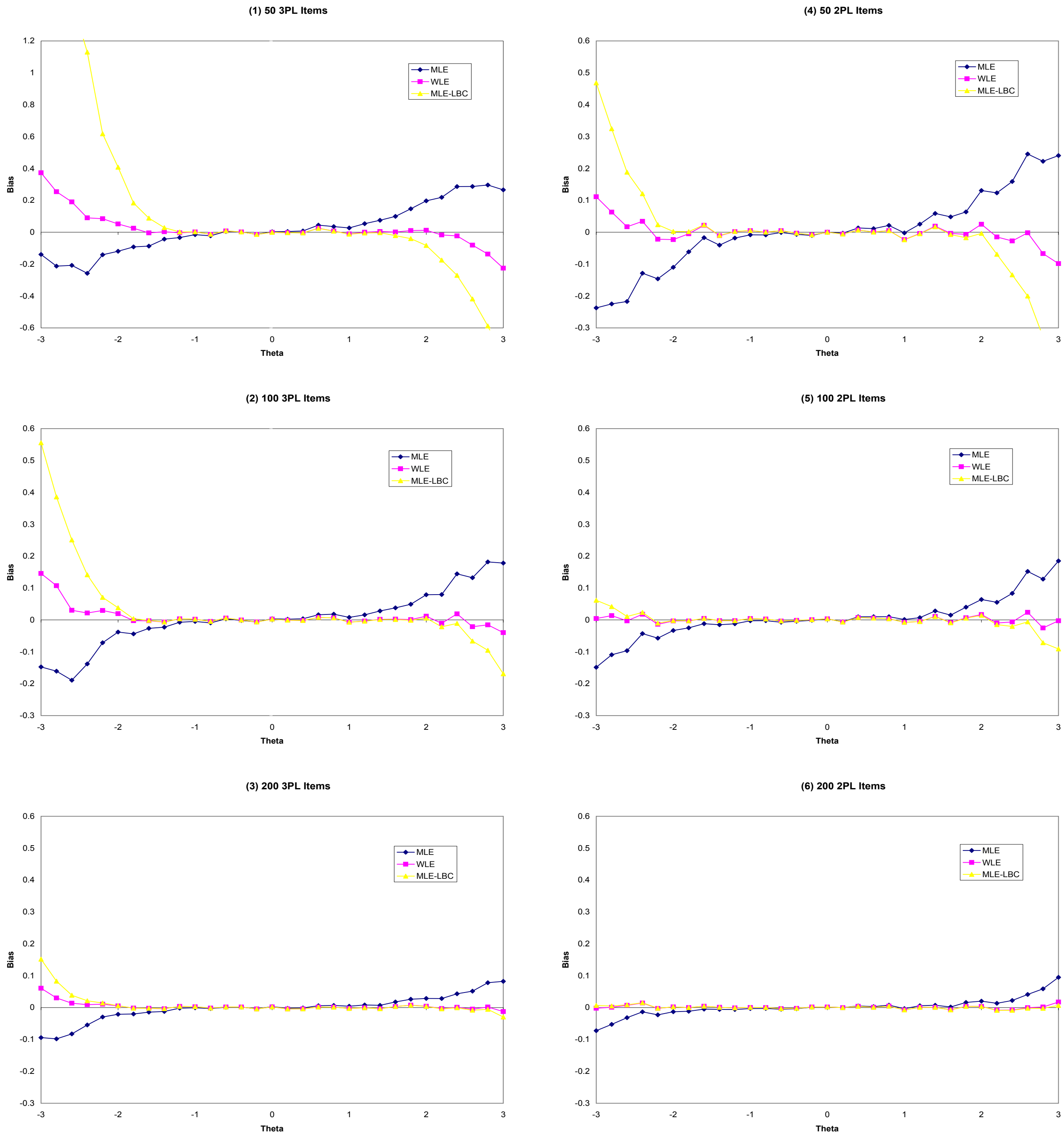

Figure 2. Bias of 3PL and 2PL ability estimates from MLE, WLE, and MLE-LBC when the true item parameters are known.

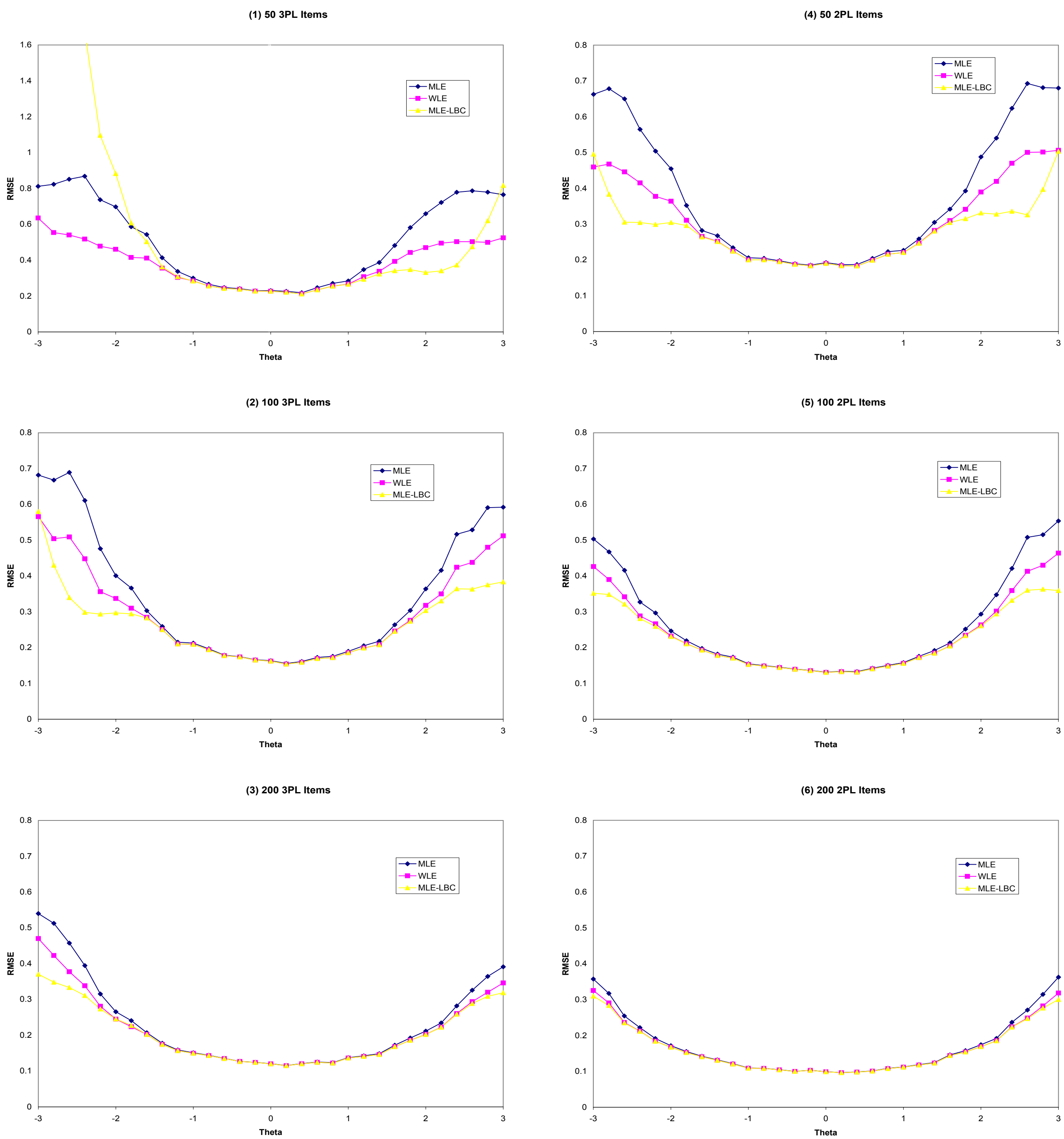

Figure 3. **RMSE of 3PL and 2PL ability estimates from MLE, WLE, and MLE-LBC when the true item parameters are known.**

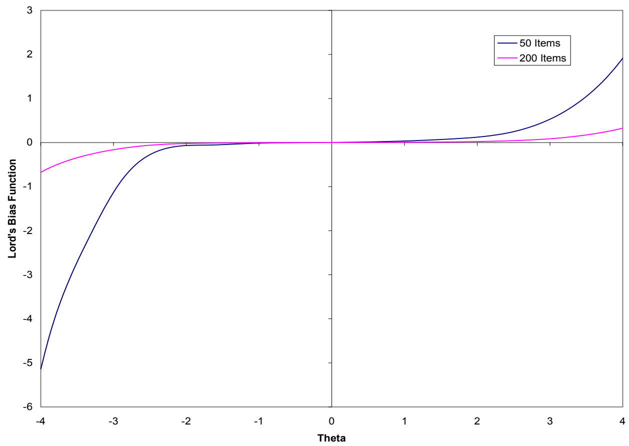

Figure 4. **Lord's bias functions for 50 and 200 3PL items.**

When the estimated item parameters from PARSCALE are used as substitutes for true item parameters to estimate abilities in the 2PL cases, the pattern of the bias of the MLE remains the same as that when the item parameters are given (see Figures 2, 5, and 7). Hence, WLE in the 2PL cases and MLE-LBC in the cases of 100 or 200 2PL items still effectively reduce the bias of MLE when item parameters are estimated by PARSCALE. However, the pattern of the bias of the MLE in the cases of 100 or 200 3PL items is no longer like what was demonstrated by Lord (1983) when true item parameters are given (see Figures 5 and 7). Consequently, the direction in which the bias is corrected for with WLE and MLE-LBC is not correct, and the average bias of WLE and MLE-LBC is worse (larger) than that of MLE at most ability levels under these conditions. The effect of bias correction is mixed at different ability levels for the case of 50 3PL items; that is, the average bias of WLE at some levels is smaller than that of MLE, but larger at the other levels. Overall, neither MLE-LBC nor WLE very successfully reduces the bias of MLE in the 3PL cases in the simulation studies when item parameters are estimated. Figures 6 and 8 show that WLE and MLE-LBC in the cases of 100 and 200 items have approximately the same (or slightly smaller) RMSE as MLE. It is important to note that when item parameters are estimated using a calibration sample, the effectiveness of ability bias correction depends on the program (software) used to estimate item parameters.

Table 1

Sample Correlations Between Estimated and True Item Parameters and Average RMSE of Estimated IRF From Calibration Samples Using PARSCALE

Model	Number of examinees	Number of items	Correlations *a*	*b*	*c*	Ave (SD) of RMSE of estimated IRF
3PL	500	50	0.8498	0.9695	0.7815	0.0283 (0.0119)
		100	0.8463	0.9688	0.7108	0.0334 (0.0153)
		200	0.8407	0.9659	0.6848	0.0372 (0.0175)
	1,000	50	0.8783	0.9752	0.8103	0.0206 (0.0087)
		100	0.8712	0.9747	0.7561	0.0264 (0.0117)
		200	0.8876	0.9772	0.7407	0.0351 (0.0151)
2PL	500	50	0.9419	0.9960	NA	0.0214 (0.0105)
		100	0.9330	0.9940	NA	0.0228 (0.0118)
		200	0.9402	0.9949	NA	0.0224 (0.0116)
	1,000	50	0.9781	0.9979	NA	0.0135 (0.0079)
		100	0.9752	0.9969	NA	0.0167 (0.0090)
		200	0.9734	0.9973	NA	0.0168 (0.0087)

Table 2

Average Bias and RMSE Between Estimated and True Item Parameters From Calibration Samples Using PARSCALE

Model	Number of examinees	Number of items	Ave Bias *a*	Ave Bias *b*	Ave Bias *c*	RMSEs *a*	RMSEs *b*	RMSEs *c*
3PL	500	50	0.0390	-0.0763	-0.0191	0.2540	0.2299	0.0814
		100	0.0535	-0.1243	-0.0238	0.2416	0.2695	0.0913
		200	0.0542	-0.1393	-0.0071	0.2390	0.3023	0.0849
	1,000	50	0.0500	-0.0382	0.0005	0.2274	0.2009	0.0700
		100	0.0480	-0.1074	-0.0069	0.2264	0.2455	0.0802
		200	0.0379	-0.1505	-0.0008	0.1968	0.2660	0.0777
2PL	500	50	0.0177	0.0054	NA	0.1359	0.0789	NA
		100	-0.0005	0.0020	NA	0.1393	0.1058	NA
		200	-0.0347	0.0098	NA	0.1302	0.1120	NA
	1,000	50	-0.0130	-0.0154	NA	0.0790	0.0710	NA
		100	-0.0305	-0.0169	NA	0.0888	0.0935	NA
		200	-0.0413	-0.0052	NA	0.0935	0.0911	NA

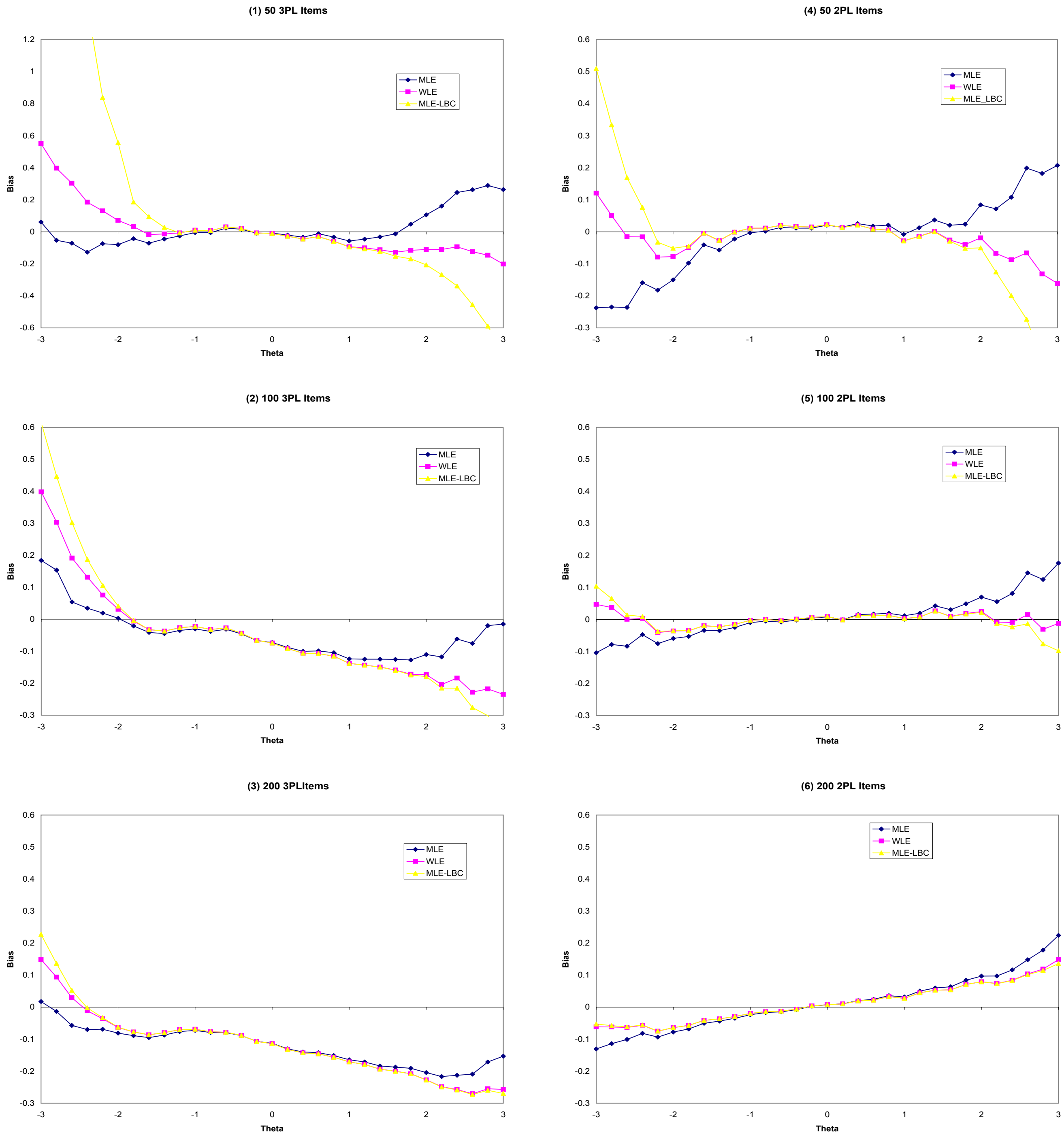

Figure 5. Bias of 3PL and 2PL ability estimates from MLE, WLE, and MLE-LBC with estimated item parameters from PARSCALE (calibration sample size 500).

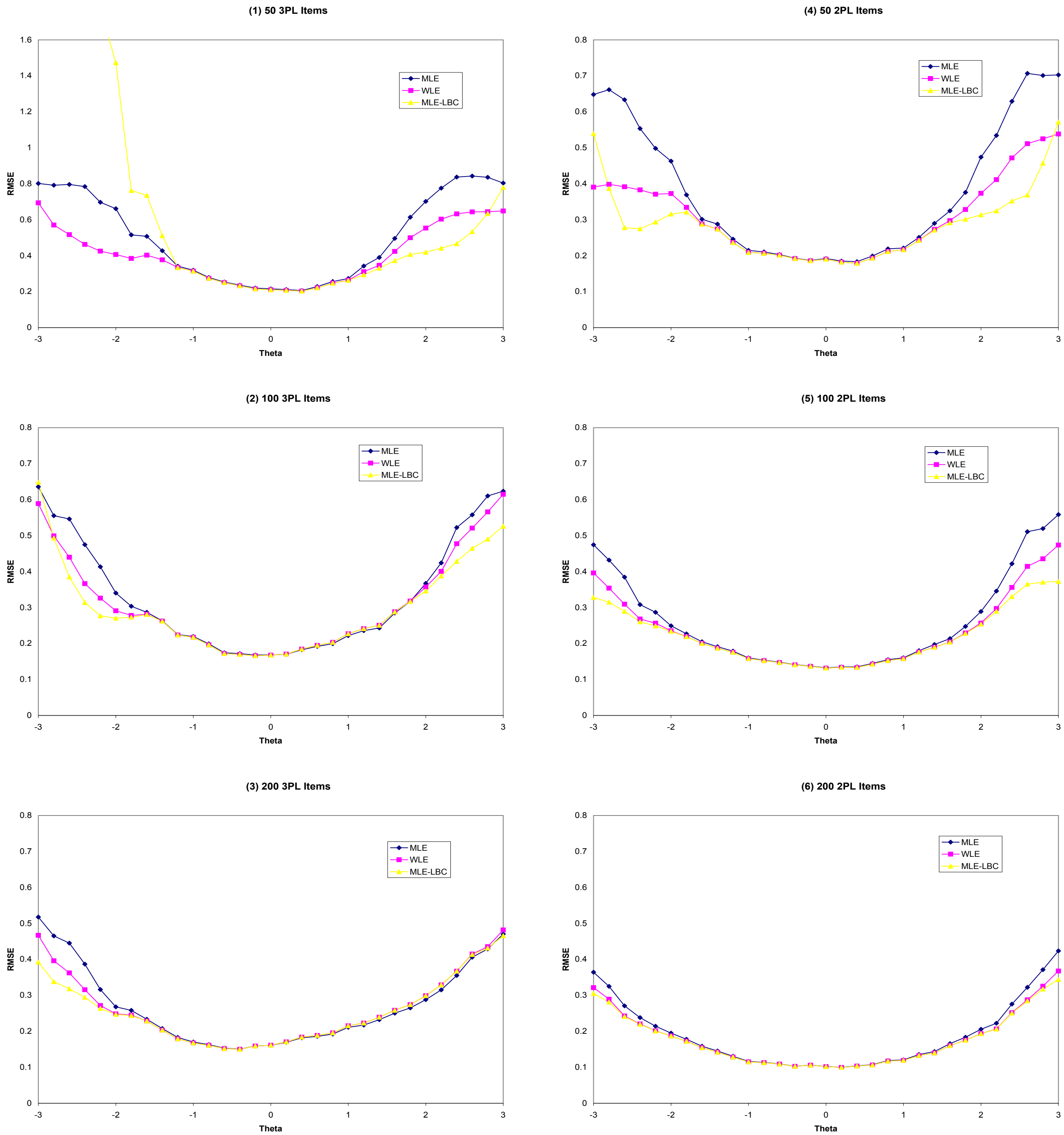

Figure 6. RMSE of 3PL and 2PL ability estimates from MLE, WLE, and MLE-LBC with estimated item parameters from PARSCALE (calibration sample size 500).

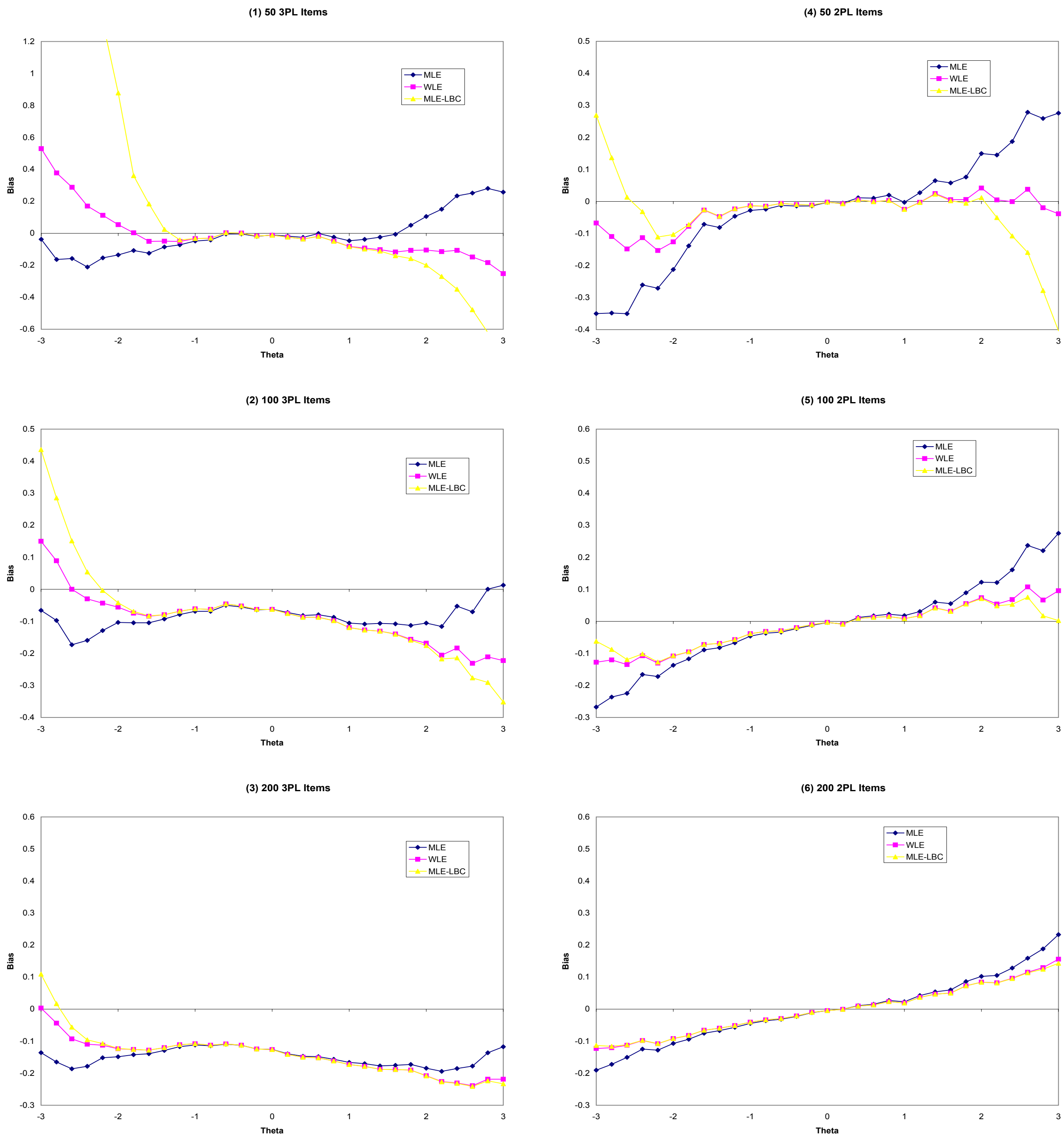

Figure 7. Bias of 3PL and 2PL ability estimates from MLE, WLE, and MLE-LBC with estimated item parameters from PARSCALE (calibration sample size 1,000).

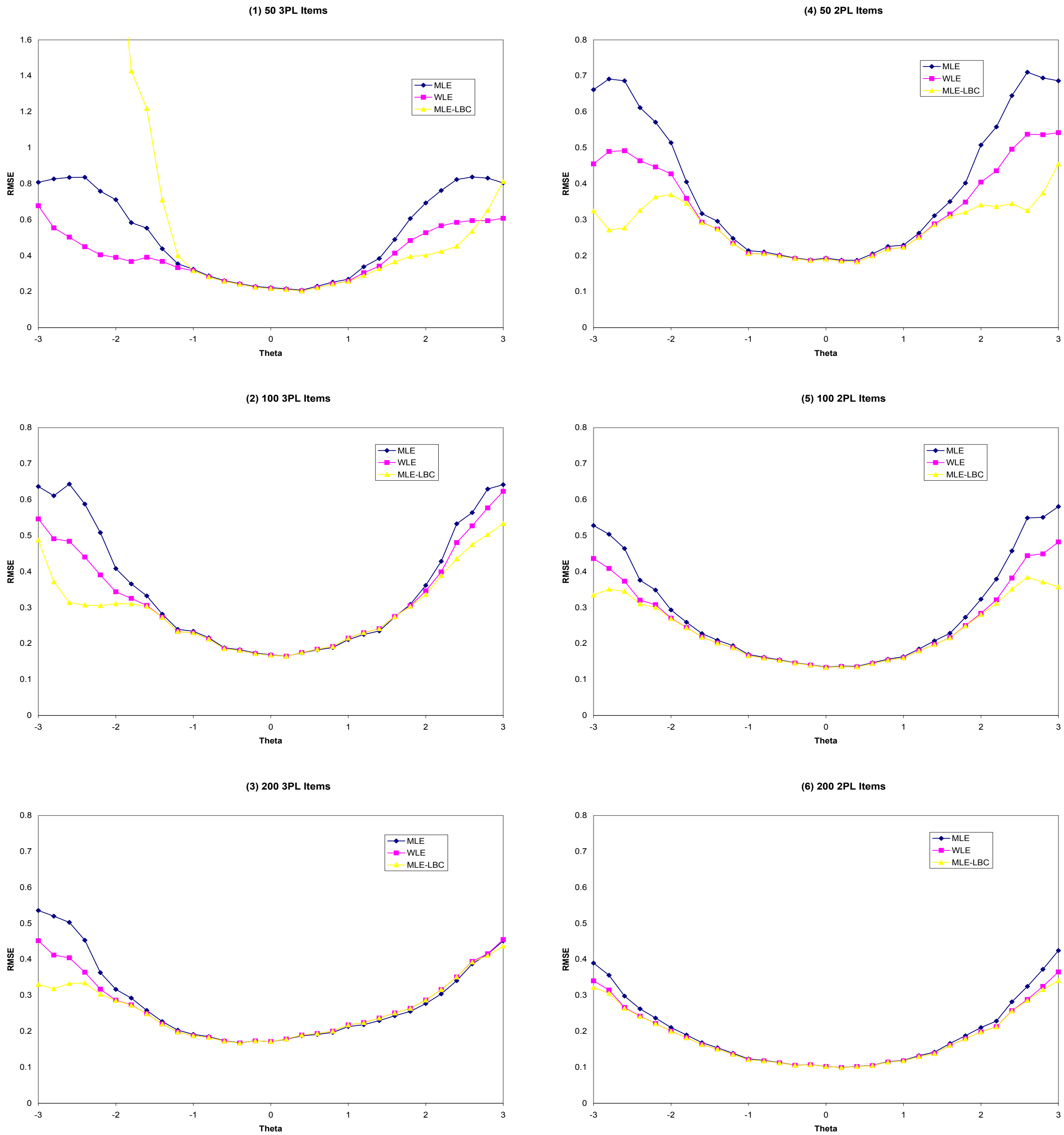

Figure 8. **RMSE of 3PL and 2PL ability estimates from MLE, WLE, and MLE-LBC with estimated item parameters from PARSCALE (calibration sample size 1,000).**

When comparing the results between the cases where the true item parameters are known (Tables A1-A6) and the cases where they are unknown (Tables A7-A18), overall the RMSEs of θ increased for the cases with unknown item parameters but not by very much. Note that the design of a calibration sample in this simulation was in favor of getting accurate parameter estimates. In fact, the calibration samples used in this paper were "perfectly" generated in the sense that there was no systematic statistical noise added into the process of response data generation. The standard normal distribution used to generate the true abilities for a calibration sample was even applied as a prior distribution in estimating item parameters by PARSCALE.

4. Discussion

Based on the results of these simulation studies, WLE and MLE-LBC are effective in reducing the bias of MLE in the cases when the true item parameters are known. However, when item parameters are unknown, as would be the case in most applications, these two methods are not as effective for the 3PL cases. If estimated item parameters are treated as fixed, without statistical errors, and are used as substitutes for the true item parameters in estimating abilities, the MLE $\hat{\theta}$ based on these fixed estimated item parameters will converge to some value, say θ^*, according to large sample theory under proper regularity conditions, when the number of items becomes larger and larger. However, θ^* will not necessarily be the true ability parameter θ. What WLE and MLE-LBC try to do is to reduce the "bias" against θ^*, not the bias against the true θ. The bias of the MLE of θ based on fixed estimated item parameters comes from at least two sources: the statistical error of MLE given item parameters and the statistical error from the item parameter estimates. The former source was investigated by Lord (1983, 1986), Warm (1989), and Samejima (1993a, 1993b). To obtain a better bias-adjusted MLE of an ability parameter in most practical situations where item parameters are unknown, both error sources should be considered simultaneously. Specifically, estimated item parameters should be regarded as covariates measured with error instead of treating them as fixed. Thus, a bias correction formula can be developed along the line of what has been done in research on measurement error models (Stefanski & Carroll, 1985; Song, 2003). This research is under way and its results, including the comparison of this new method with WLE and MLE-LBC, will be reported in a separate paper.

The accuracy of ability estimates is very important since estimated ability scores are the major measurement output of a test that is analyzed using IRT. The results of this study are useful for CBT situations, where item parameters are based on calibration done previously with a calibration sample and the resulting estimates are assumed to be fixed. The results also shed some light on statistical inference based on estimated item parameters being used with target response data.

References

Allen, N., Carlson, J. E., & Zelenak, C. (1999). *The NAEP 1996 technical report* (NCES 1999-452). Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Birnbaum, A. (1968). Some latent ability models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 392-479). Reading, MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of the EM algorithm. *Psychometrika, 46*, 443-459.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika, 48*, 233-245.

Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement, 23*, 157-162.

Mislevy, R., & Bock, R. D. (1982). BILOG: Item analysis and test scoring with binary logistic models [Computer software]. Mooresville, IN: Scientific Software, Inc.

Muraki, E., & Bock, R. D. (1991). PARSCALE: Parameter scaling of rating data [Computer software]. Chicago, IL: Scientific Software, Inc.

Muraki, E., & Bock, R. D. (1997). PARSCALE: IRT item analysis and test scoring for rating scale data [Computer software]. Chicago, IL: Scientific Software, International.

Samejima, F. (1993a). An approximation for the bias function of the maximum likelihood estimate of a latent variable for the general case where the item responses are discrete. *Psychometrika, 58*, 119-138.

Samejima, F. (1993b). The bias function of the maximum likelihood estimate of ability for dichotomous response level. *Psychometrika, 58*, 195-209.

Song, X. (2003). *Item parameter measurement error in item response theory models.* Unpublished doctoral dissertation, Department of Statistics, Rutgers, The State University of New Jersey.

Stefanski, L. A., & Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Annals of Statistics, 13*, 1335-1351.

Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement, 35*, 109-135.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427-450.

Yi, Q., Wang, T., & Ban, J. (2001). Effect of scale transformation and test-termination rule on the precision of ability estimation in computerized adaptive testing. *Journal of Educational Measurement, 38*, 267-292.

Appendix

Table A1

Average Bias and RMSE of Ability Estimates With 50 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Known

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.00	−0.1391	0.3730	2.3016	0.8120	0.6353	2.7255
−2.80	−0.2126	0.2540	1.8659	0.8235	0.5542	2.3311
−2.60	−0.2078	0.1904	1.4532	0.8514	0.5408	1.9640
−2.40	−0.2577	0.0903	1.1302	0.8680	0.5177	1.6777
−2.20	−0.1411	0.0847	0.6177	0.7373	0.4780	1.0964
−2.00	−0.1192	0.0521	0.4080	0.6976	0.4609	0.8827
−1.80	−0.0921	0.0250	0.1838	0.5865	0.4152	0.6078
−1.60	−0.0872	−0.0038	0.0887	0.5435	0.4119	0.5034
−1.40	−0.0433	0.0034	0.0291	0.4138	0.3559	0.3624
−1.20	−0.0335	−0.0034	0.0018	0.3377	0.3041	0.3083
−1.00	−0.0151	0.0019	0.0026	0.3001	0.2847	0.2840
−0.80	−0.0222	−0.0131	−0.0127	0.2660	0.2566	0.2562
−0.60	0.0049	0.0084	0.0086	0.2475	0.2430	0.2428
−0.40	0.0014	0.0023	0.0023	0.2413	0.2385	0.2385
−0.20	−0.0126	−0.0137	−0.0137	0.2297	0.2274	0.2274
0.00	0.0030	−0.0006	−0.0007	0.2306	0.2272	0.2272
0.20	0.0045	−0.0027	−0.0028	0.2262	0.2216	0.2215
0.40	0.0080	−0.0039	−0.0042	0.2186	0.2122	0.2120
0.60	0.0449	0.0250	0.0244	0.2471	0.2346	0.2343
0.80	0.0357	0.0078	0.0067	0.2708	0.2558	0.2550
1.00	0.0273	−0.0104	−0.0119	0.2845	0.2671	0.2661
1.20	0.0536	0.0000	−0.0049	0.3477	0.3084	0.2936
1.40	0.0750	0.0043	−0.0028	0.3875	0.3379	0.3229
1.60	0.0993	0.0016	−0.0209	0.4822	0.3933	0.3413
1.80	0.1473	0.0107	−0.0401	0.5815	0.4433	0.3479
2.00	0.1972	0.0123	−0.0827	0.6593	0.4699	0.3315
2.20	0.2191	−0.0171	−0.1744	0.7218	0.4951	0.3405
2.40	0.2865	−0.0227	−0.2701	0.7787	0.5033	0.3732
2.60	0.2873	−0.0807	−0.4187	0.7868	0.5032	0.4749
2.80	0.2963	−0.1370	−0.5881	0.7796	0.4993	0.6200
3.00	0.2656	−0.2252	−0.7961	0.7651	0.5249	0.8167

Table A2

Average Bias and RMSE of Ability Estimates With 100 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Known

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.00	−0.1473	0.1457	0.5555	0.6818	0.5659	0.5804
−2.80	−0.1607	0.1073	0.3859	0.6678	0.5043	0.4301
−2.60	−0.1890	0.0303	0.2509	0.6893	0.5092	0.3397
−2.40	−0.1378	0.0217	0.1416	0.6109	0.4481	0.2984
−2.20	−0.0714	0.0296	0.0708	0.4763	0.3563	0.2936
−2.00	−0.0379	0.0197	0.0381	0.4007	0.3373	0.2969
−1.80	−0.0437	−0.0022	0.0039	0.3663	0.3101	0.2943
−1.60	−0.0263	−0.0025	−0.0014	0.3031	0.2847	0.2829
−1.40	−0.0226	−0.0067	−0.0064	0.2590	0.2504	0.2503
−1.20	−0.0072	0.0036	0.0037	0.2152	0.2102	0.2104
−1.00	−0.0049	0.0022	0.0023	0.2126	0.2090	0.2091
−0.80	−0.0093	−0.0049	−0.0049	0.1967	0.1942	0.1942
−0.60	0.0034	0.0059	0.0059	0.1788	0.1776	0.1776
−0.40	−0.0016	−0.0003	−0.0003	0.1748	0.1739	0.1739
−0.20	−0.0064	−0.0062	−0.0062	0.1662	0.1652	0.1652
0.00	0.0038	0.0025	0.0025	0.1636	0.1622	0.1622
0.20	0.0023	−0.0008	−0.0008	0.1557	0.1541	0.1541
0.40	0.0037	−0.0017	−0.0018	0.1611	0.1590	0.1590
0.60	0.0161	0.0077	0.0076	0.1725	0.1693	0.1693
0.80	0.0181	0.0065	0.0064	0.1757	0.1720	0.1719
1.00	0.0080	−0.0070	−0.0072	0.1897	0.1859	0.1858
1.20	0.0159	−0.0037	−0.0040	0.2052	0.1990	0.1987
1.40	0.0280	0.0020	0.0014	0.2179	0.2080	0.2077
1.60	0.0380	0.0026	0.0014	0.2637	0.2464	0.2453
1.80	0.0493	0.0010	−0.0012	0.3038	0.2762	0.2736
2.00	0.0789	0.0118	0.0055	0.3640	0.3180	0.3035
2.20	0.0797	−0.0106	−0.0214	0.4158	0.3500	0.3303
2.40	0.1444	0.0190	−0.0108	0.5166	0.4245	0.3642
2.60	0.1323	−0.0211	−0.0662	0.5288	0.4380	0.3633
2.80	0.1821	−0.0155	−0.0953	0.5909	0.4802	0.3752
3.00	0.1782	−0.0398	−0.1689	0.5921	0.5122	0.3835

Table A3

Average Bias and RMSE of Ability Estimates With 200 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Known

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.00	−0.0940	0.0607	0.1518	0.5399	0.4702	0.3705
−2.80	−0.0981	0.0303	0.0830	0.5127	0.4229	0.3482
−2.60	−0.0826	0.0138	0.0387	0.4575	0.3776	0.3334
−2.40	−0.0543	0.0093	0.0215	0.3947	0.3384	0.3111
−2.20	−0.0293	0.0107	0.0145	0.3154	0.2811	0.2735
−2.00	−0.0210	0.0052	0.0059	0.2656	0.2452	0.2455
−1.80	−0.0199	−0.0013	−0.0016	0.2410	0.2240	0.2272
−1.60	−0.0142	−0.0020	−0.0019	0.2071	0.2023	0.2023
−1.40	−0.0120	−0.0035	−0.0034	0.1777	0.1745	0.1745
−1.20	−0.0016	0.0038	0.0039	0.1590	0.1569	0.1569
−1.00	−0.0006	0.0026	0.0027	0.1512	0.1498	0.1498
−0.80	−0.0030	−0.0013	−0.0013	0.1441	0.1431	0.1432
−0.60	0.0011	0.0018	0.0019	0.1355	0.1350	0.1350
−0.40	0.0017	0.0018	0.0018	0.1271	0.1268	0.1268
−0.20	−0.0040	−0.0043	−0.0043	0.1244	0.1241	0.1241
0.00	0.0031	0.0023	0.0023	0.1207	0.1202	0.1202
0.20	−0.0024	−0.0041	−0.0041	0.1155	0.1150	0.1150
0.40	−0.0012	−0.0038	−0.0039	0.1210	0.1204	0.1204
0.60	0.0059	0.0019	0.0018	0.1254	0.1244	0.1244
0.80	0.0069	0.0015	0.0014	0.1235	0.1224	0.1224
1.00	0.0044	−0.0025	−0.0026	0.1376	0.1365	0.1365
1.20	0.0082	−0.0005	−0.0005	0.1426	0.1410	0.1410
1.40	0.0072	−0.0036	−0.0037	0.1485	0.1464	0.1463
1.60	0.0178	0.0036	0.0035	0.1723	0.1682	0.1682
1.80	0.0264	0.0077	0.0075	0.1925	0.1860	0.1859
2.00	0.0286	0.0044	0.0039	0.2116	0.2026	0.2023
2.20	0.0284	−0.0030	−0.0038	0.2345	0.2229	0.2223
2.40	0.0434	0.0010	−0.0005	0.2821	0.2606	0.2588
2.60	0.0517	−0.0048	−0.0082	0.3258	0.2940	0.2887
2.80	0.0780	0.0020	−0.0053	0.3644	0.3202	0.3088
3.00	0.0825	−0.0124	−0.0295	0.3912	0.3462	0.3187

Table A4

Average Bias and RMSE of Ability Estimates With 50 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Known

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.00	−0.2373	0.1113	0.4683	0.6626	0.4596	0.4954
−2.80	−0.2246	0.0632	0.3245	0.6783	0.4678	0.3834
−2.60	−0.2171	0.0171	0.1884	0.6499	0.4461	0.3055
−2.40	−0.1286	0.0341	0.1207	0.5648	0.4153	0.3046
−2.20	−0.1462	−0.0218	0.0236	0.5040	0.3776	0.2991
−2.00	−0.1099	−0.0231	0.0025	0.4546	0.3641	0.3046
−1.80	−0.0613	−0.0047	0.0019	0.3522	0.3108	0.2961
−1.60	−0.0169	0.0216	0.0232	0.2822	0.2658	0.2643
−1.40	−0.0403	−0.0110	−0.0099	0.2676	0.2520	0.2513
−1.20	−0.0182	0.0018	0.0024	0.2342	0.2247	0.2244
−1.00	−0.0081	0.0051	0.0054	0.2061	0.2003	0.2001
−0.80	−0.0084	0.0004	0.0005	0.2046	0.2005	0.2004
−0.60	−0.0007	0.0049	0.0050	0.1978	0.1954	0.1954
−0.40	−0.0066	−0.0030	−0.0029	0.1894	0.1877	0.1877
−0.20	−0.0099	−0.0080	−0.0080	0.1855	0.1837	0.1837
0.00	0.0004	0.0002	0.0002	0.1925	0.1903	0.1903
0.20	−0.0029	−0.0057	−0.0058	0.1866	0.1839	0.1839
0.40	0.0134	0.0070	0.0069	0.1872	0.1832	0.1831
0.60	0.0114	0.0007	0.0005	0.2043	0.1992	0.1991
0.80	0.0214	0.0054	0.0050	0.2232	0.2160	0.2158
1.00	−0.0019	−0.0231	−0.0236	0.2268	0.2207	0.2206
1.20	0.0254	−0.0039	−0.0047	0.2587	0.2471	0.2466
1.40	0.0590	0.0189	0.0173	0.3050	0.2825	0.2804
1.60	0.0483	−0.0037	−0.0071	0.3419	0.3101	0.3045
1.80	0.0637	−0.0065	−0.0174	0.3928	0.3414	0.3153
2.00	0.1307	0.0252	−0.0035	0.4877	0.3897	0.3311
2.20	0.1234	−0.0147	−0.0690	0.5403	0.4195	0.3279
2.40	0.1590	−0.0273	−0.1338	0.6235	0.4701	0.3360
2.60	0.2453	−0.0016	−0.2001	0.6928	0.5005	0.3258
2.80	0.2225	−0.0668	−0.3283	0.6813	0.5014	0.3968
3.00	0.2405	−0.0982	−0.4700	0.6801	0.5063	0.5034

Table A5

Average Bias and RMSE of Ability Estimates With 100 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Known

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.00	−0.1485	0.0045	0.0617	0.5033	0.4264	0.3518
−2.80	−0.1094	0.0136	0.0419	0.4673	0.3901	0.3484
−2.60	−0.0965	−0.0028	0.0108	0.4160	0.3419	0.3214
−2.40	−0.0428	0.0179	0.0236	0.3272	0.2884	0.2811
−2.20	−0.0572	−0.0135	−0.0101	0.2968	0.2664	0.2593
−2.00	−0.0330	−0.0033	−0.0021	0.2465	0.2322	0.2310
−1.80	−0.0250	−0.0031	−0.0026	0.2192	0.2110	0.2105
−1.60	−0.0119	0.0047	0.0049	0.1977	0.1931	0.1930
−1.40	−0.0150	−0.0022	−0.0020	0.1819	0.1781	0.1781
−1.20	−0.0121	−0.0024	−0.0023	0.1734	0.1705	0.1705
−1.00	−0.0025	0.0045	0.0045	0.1547	0.1530	0.1530
−0.80	−0.0024	0.0027	0.0027	0.1498	0.1486	0.1486
−0.60	−0.0072	−0.0035	−0.0034	0.1453	0.1443	0.1443
−0.40	−0.0042	−0.0015	−0.0015	0.1401	0.1394	0.1394
−0.20	−0.0011	0.0006	0.0006	0.1362	0.1355	0.1355
0.00	0.0031	0.0036	0.0036	0.1316	0.1308	0.1308
0.20	−0.0058	−0.0066	−0.0066	0.1337	0.1328	0.1328
0.40	0.0099	0.0073	0.0073	0.1329	0.1315	0.1315
0.60	0.0103	0.0059	0.0059	0.1425	0.1408	0.1408
0.80	0.0102	0.0038	0.0037	0.1505	0.1486	0.1486
1.00	0.0013	−0.0075	−0.0075	0.1580	0.1560	0.1559
1.20	0.0071	−0.0050	−0.0052	0.1757	0.1722	0.1721
1.40	0.0281	0.0110	0.0106	0.1920	0.1848	0.1845
1.60	0.0149	−0.0082	−0.0089	0.2131	0.2047	0.2043
1.80	0.0399	0.0068	0.0056	0.2516	0.2346	0.2335
2.00	0.0640	0.0168	0.0143	0.2934	0.2635	0.2607
2.20	0.0551	−0.0097	−0.0149	0.3476	0.3020	0.2941
2.40	0.0834	−0.0064	−0.0202	0.4211	0.3594	0.3320
2.60	0.1521	0.0240	−0.0060	0.5080	0.4131	0.3602
2.80	0.1280	−0.0251	−0.0714	0.5152	0.4301	0.3631
3.00	0.1850	−0.0026	−0.0909	0.5538	0.4640	0.3594

Table A6

Average Bias and RMSE of Ability Estimates With 200 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Known

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.00	−0.0724	−0.0020	0.0068	0.3576	0.3254	0.3097
−2.80	−0.0526	0.0010	0.0050	0.3174	0.2908	0.2837
−2.60	−0.0317	0.0071	0.0086	0.2547	0.2364	0.2353
−2.40	−0.0132	0.0144	0.0152	0.2222	0.2123	0.2118
−2.20	−0.0229	−0.0025	−0.0021	0.1918	0.1843	0.1840
−2.00	−0.0130	0.0021	0.0023	0.1716	0.1675	0.1674
−1.80	−0.0115	0.0001	0.0002	0.1554	0.1527	0.1526
−1.60	−0.0047	0.0043	0.0044	0.1423	0.1407	0.1407
−1.40	−0.0058	0.0010	0.0011	0.1322	0.1307	0.1307
−1.20	−0.0059	−0.0009	−0.0008	0.1215	0.1204	0.1204
−1.00	−0.0030	0.0005	0.0006	0.1096	0.1089	0.1089
−0.80	−0.0025	0.0001	0.0001	0.1087	0.1082	0.1082
−0.60	−0.0052	−0.0034	−0.0034	0.1048	0.1044	0.1044
−0.40	−0.0036	−0.0023	−0.0023	0.1002	0.0999	0.0999
−0.20	0.0012	0.0020	0.0020	0.1033	0.1030	0.1030
0.00	0.0017	0.0019	0.0019	0.0992	0.0989	0.0989
0.20	0.0002	−0.0002	−0.0002	0.0966	0.0963	0.0963
0.40	0.0052	0.0041	0.0041	0.0986	0.0981	0.0981
0.60	0.0031	0.0011	0.0011	0.1012	0.1007	0.1007
0.80	0.0076	0.0047	0.0047	0.1085	0.1079	0.1079
1.00	−0.0034	−0.0072	−0.0072	0.1123	0.1119	0.1119
1.20	0.0059	0.0009	0.0009	0.1189	0.1179	0.1179
1.40	0.0074	0.0007	0.0007	0.1246	0.1231	0.1231
1.60	0.0020	−0.0070	−0.0071	0.1459	0.1440	0.1440
1.80	0.0163	0.0040	0.0039	0.1578	0.1541	0.1541
2.00	0.0203	0.0038	0.0036	0.1744	0.1690	0.1689
2.20	0.0135	−0.0082	−0.0086	0.1920	0.1855	0.1853
2.40	0.0224	−0.0075	−0.0083	0.2371	0.2242	0.2233
2.60	0.0413	−0.0001	−0.0019	0.2711	0.2490	0.2472
2.80	0.0588	0.0018	−0.0022	0.3146	0.2825	0.2765
3.00	0.0950	0.0176	0.0069	0.3627	0.3185	0.3003

Table A7

Average Bias and RMSE of Ability Estimates With 50 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 500

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	0.0615	0.5507	2.8600	0.8016	0.6941	3.8152
−2.8	−0.0525	0.3974	2.4202	0.7920	0.5707	3.4229
−2.6	−0.0709	0.3035	1.8671	0.7962	0.5172	2.9042
−2.4	−0.1266	0.1851	1.4170	0.7844	0.4628	2.4729
−2.2	−0.0739	0.1308	0.8393	0.6970	0.4254	1.7719
−2.0	−0.0798	0.0715	0.5569	0.6614	0.4067	1.4733
−1.8	−0.0430	0.0322	0.1865	0.5157	0.3848	0.7628
−1.6	−0.0712	−0.0171	0.0946	0.5077	0.4032	0.7351
−1.4	−0.0446	−0.0131	0.0277	0.4277	0.3772	0.5114
−1.2	−0.0257	−0.0068	−0.0068	0.3414	0.3350	0.3336
−1.0	−0.0051	0.0106	0.0105	0.3193	0.3141	0.3136
−0.8	−0.0052	0.0070	0.0069	0.2787	0.2739	0.2743
−0.6	0.0228	0.0311	0.0310	0.2538	0.2502	0.2505
−0.4	0.0164	0.0214	0.0214	0.2361	0.2323	0.2324
−0.2	−0.0083	−0.0065	−0.0066	0.2202	0.2156	0.2157
0.0	−0.0078	−0.0105	−0.0106	0.2155	0.2107	0.2106
0.2	−0.0202	−0.0272	−0.0273	0.2117	0.2076	0.2076
0.4	−0.0334	−0.0449	−0.0450	0.2062	0.2033	0.2033
0.6	−0.0117	−0.0299	−0.0302	0.2284	0.2215	0.2212
0.8	−0.0329	−0.0587	−0.0596	0.2563	0.2478	0.2468
1.0	−0.0568	−0.0929	−0.0947	0.2732	0.2639	0.2627
1.2	−0.0453	−0.1004	−0.1070	0.3427	0.3109	0.2936
1.4	−0.0315	−0.1121	−0.1217	0.3901	0.3466	0.3308
1.6	−0.0125	−0.1273	−0.1524	0.4960	0.4239	0.3727
1.8	0.0480	−0.1155	−0.1688	0.6139	0.4999	0.4069
2.0	0.1059	−0.1107	−0.2065	0.7018	0.5536	0.4194
2.2	0.1609	−0.1103	−0.2678	0.7758	0.6033	0.4417
2.4	0.2454	−0.0932	−0.3381	0.8373	0.6325	0.4666
2.6	0.2626	−0.1235	−0.4555	0.8430	0.6437	0.5340
2.8	0.2894	−0.1460	−0.5883	0.8360	0.6447	0.6347
3.0	0.2636	−0.2013	−0.7539	0.8037	0.6488	0.7794

Table A8

Average Bias and RMSE of Ability Estimates With 100 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 500

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	0.1842	0.3987	0.6152	0.6358	0.5890	0.6480
−2.8	0.1540	0.3038	0.4472	0.5553	0.4993	0.4932
−2.6	0.0543	0.1916	0.3024	0.5463	0.4399	0.3849
−2.4	0.0348	0.1319	0.1869	0.4749	0.3665	0.3139
−2.2	0.0196	0.0760	0.1057	0.4134	0.3263	0.2770
−2.0	0.0032	0.0316	0.0414	0.3403	0.2910	0.2705
−1.8	−0.0208	−0.0058	−0.0036	0.3036	0.2782	0.2730
−1.6	−0.0411	−0.0323	−0.0321	0.2868	0.2814	0.2806
−1.4	−0.0447	−0.0367	−0.0369	0.2631	0.2623	0.2621
−1.2	−0.0348	−0.0263	−0.0266	0.2248	0.2236	0.2238
−1.0	−0.0299	−0.0222	−0.0223	0.2195	0.2174	0.2175
−0.8	−0.0380	−0.0318	−0.0319	0.1991	0.1962	0.1963
−0.6	−0.0311	−0.0269	−0.0269	0.1743	0.1720	0.1720
−0.4	−0.0458	−0.0434	−0.0434	0.1719	0.1697	0.1697
−0.2	−0.0663	−0.0656	−0.0656	0.1680	0.1662	0.1662
0.0	−0.0727	−0.0742	−0.0742	0.1685	0.1676	0.1676
0.2	−0.0879	−0.0916	−0.0916	0.1701	0.1706	0.1706
0.4	−0.1001	−0.1060	−0.1060	0.1827	0.1845	0.1845
0.6	−0.0989	−0.1073	−0.1072	0.1921	0.1948	0.1948
0.8	−0.1044	−0.1154	−0.1153	0.1994	0.2033	0.2032
1.0	−0.1239	−0.1379	−0.1379	0.2220	0.2272	0.2271
1.2	−0.1247	−0.1433	−0.1434	0.2357	0.2414	0.2413
1.4	−0.1247	−0.1492	−0.1496	0.2432	0.2506	0.2506
1.6	−0.1254	−0.1586	−0.1596	0.2848	0.2883	0.2877
1.8	−0.1272	−0.1719	−0.1740	0.3170	0.3182	0.3171
2.0	−0.1102	−0.1728	−0.1794	0.3672	0.3572	0.3460
2.2	−0.1177	−0.2041	−0.2150	0.4242	0.4006	0.3880
2.4	−0.0615	−0.1837	−0.2155	0.5224	0.4775	0.4284
2.6	−0.0756	−0.2278	−0.2759	0.5578	0.5209	0.4648
2.8	−0.0198	−0.2177	−0.3015	0.6103	0.5659	0.4901
3.0	−0.0149	−0.2348	−0.3701	0.6235	0.6149	0.5260

Table A9

Average Bias and RMSE of Ability Estimates With 200 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 500

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	0.0176	0.1490	0.2276	0.5176	0.4670	0.3924
−2.8	−0.0133	0.0941	0.1364	0.4652	0.3962	0.3381
−2.6	−0.0570	0.0294	0.0525	0.4452	0.3624	0.3183
−2.4	−0.0698	−0.0113	−0.0020	0.3869	0.3157	0.2945
−2.2	−0.0690	−0.0360	−0.0327	0.3162	0.2718	0.2635
−2.0	−0.0811	−0.0633	−0.0625	0.2680	0.2485	0.2472
−1.8	−0.0889	−0.0772	−0.0770	0.2583	0.2453	0.2438
−1.6	−0.0948	−0.0860	−0.0862	0.2336	0.2287	0.2287
−1.4	−0.0870	−0.0794	−0.0796	0.2078	0.2035	0.2037
−1.2	−0.0760	−0.0701	−0.0703	0.1830	0.1791	0.1793
−1.0	−0.0726	−0.0688	−0.0688	0.1704	0.1675	0.1676
−0.8	−0.0791	−0.0768	−0.0768	0.1631	0.1611	0.1611
−0.6	−0.0794	−0.0783	−0.0783	0.1532	0.1521	0.1521
−0.4	−0.0885	−0.0881	−0.0881	0.1506	0.1501	0.1501
−0.2	−0.1068	−0.1070	−0.1070	0.1593	0.1591	0.1591
0.0	−0.1128	−0.1138	−0.1138	0.1610	0.1614	0.1614
0.2	−0.1302	−0.1321	−0.1321	0.1696	0.1707	0.1707
0.4	−0.1397	−0.1425	−0.1425	0.1820	0.1838	0.1838
0.6	−0.1419	−0.1457	−0.1457	0.1860	0.1886	0.1886
0.8	−0.1510	−0.1561	−0.1561	0.1923	0.1958	0.1958
1.0	−0.1642	−0.1707	−0.1707	0.2111	0.2155	0.2155
1.2	−0.1710	−0.1791	−0.1791	0.2174	0.2230	0.2230
1.4	−0.1834	−0.1935	−0.1935	0.2324	0.2394	0.2394
1.6	−0.1871	−0.2000	−0.2001	0.2505	0.2585	0.2585
1.8	−0.1908	−0.2079	−0.2081	0.2650	0.2743	0.2744
2.0	−0.2040	−0.2270	−0.2275	0.2880	0.2991	0.2993
2.2	−0.2166	−0.2480	−0.2489	0.3152	0.3292	0.3294
2.4	−0.2126	−0.2571	−0.2585	0.3553	0.3673	0.3675
2.6	−0.2091	−0.2703	−0.2728	0.4061	0.4149	0.4136
2.8	−0.1711	−0.2549	−0.2596	0.4289	0.4354	0.4315
3.0	−0.1530	−0.2565	−0.2695	0.4704	0.4818	0.4658

Table A10

Average Bias and RMSE of Ability Estimates With 50 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 500

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.2373	0.1210	0.5097	0.6482	0.3907	0.5390
−2.8	−0.2348	0.0509	0.3342	0.6615	0.3984	0.3868
−2.6	−0.2361	−0.0155	0.1692	0.6336	0.3916	0.2780
−2.4	−0.1591	−0.0158	0.0763	0.5535	0.3830	0.2748
−2.2	−0.1820	−0.0789	−0.0326	0.4983	0.3709	0.2934
−2.0	−0.1498	−0.0772	−0.0514	0.4629	0.3728	0.3159
−1.8	−0.0975	−0.0498	−0.0446	0.3690	0.3344	0.3213
−1.6	−0.0407	−0.0048	−0.0039	0.3013	0.2882	0.2872
−1.4	−0.0568	−0.0274	−0.0267	0.2879	0.2741	0.2740
−1.2	−0.0223	−0.0011	−0.0006	0.2460	0.2365	0.2363
−1.0	−0.0031	0.0112	0.0115	0.2148	0.2091	0.2089
−0.8	0.0020	0.0116	0.0118	0.2105	0.2068	0.2066
−0.6	0.0135	0.0201	0.0202	0.2032	0.2014	0.2014
−0.4	0.0115	0.0165	0.0165	0.1929	0.1919	0.1919
−0.2	0.0116	0.0152	0.0152	0.1870	0.1859	0.1859
0.0	0.0205	0.0222	0.0223	0.1921	0.1904	0.1904
0.2	0.0145	0.0137	0.0137	0.1851	0.1822	0.1821
0.4	0.0258	0.0211	0.0210	0.1835	0.1789	0.1788
0.6	0.0183	0.0088	0.0086	0.1993	0.1932	0.1931
0.8	0.0210	0.0057	0.0053	0.2190	0.2117	0.2116
1.0	−0.0079	−0.0282	−0.0286	0.2213	0.2165	0.2165
1.2	0.0128	−0.0141	−0.0147	0.2508	0.2428	0.2425
1.4	0.0370	0.0016	0.0003	0.2899	0.2733	0.2711
1.6	0.0205	−0.0255	−0.0290	0.3247	0.2976	0.2914
1.8	0.0237	−0.0398	−0.0515	0.3759	0.3284	0.3014
2.0	0.0839	−0.0188	−0.0500	0.4740	0.3735	0.3136
2.2	0.0716	−0.0674	−0.1253	0.5343	0.4115	0.3245
2.4	0.1080	−0.0871	−0.1994	0.6293	0.4719	0.3520
2.6	0.1989	−0.0659	−0.2729	0.7068	0.5115	0.3686
2.8	0.1823	−0.1314	−0.4026	0.7010	0.5249	0.4572
3.0	0.2075	−0.1610	−0.5432	0.7026	0.5383	0.5715

Table A11

Average Bias and RMSE of Ability Estimates With 100 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 500

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.1036	0.0474	0.1039	0.4745	0.3962	0.3280
−2.8	−0.0779	0.0372	0.0653	0.4319	0.3540	0.3149
−2.6	−0.0833	0.0006	0.0143	0.3846	0.3094	0.2894
−2.4	−0.0473	0.0033	0.0091	0.3081	0.2681	0.2603
−2.2	−0.0753	−0.0406	−0.0373	0.2868	0.2566	0.2490
−2.0	−0.0591	−0.0359	−0.0351	0.2492	0.2355	0.2340
−1.8	−0.0527	−0.0349	−0.0346	0.2268	0.2193	0.2189
−1.6	−0.0340	−0.0193	−0.0192	0.2049	0.2005	0.2004
−1.4	−0.0348	−0.0226	−0.0225	0.1912	0.1869	0.1869
−1.2	−0.0246	−0.0151	−0.0151	0.1788	0.1755	0.1755
−1.0	−0.0096	−0.0025	−0.0025	0.1599	0.1581	0.1580
−0.8	−0.0053	0.0001	0.0001	0.1537	0.1525	0.1525
−0.6	−0.0078	−0.0036	−0.0036	0.1480	0.1470	0.1470
−0.4	−0.0018	0.0013	0.0013	0.1413	0.1406	0.1406
−0.2	0.0047	0.0067	0.0067	0.1372	0.1365	0.1365
0.0	0.0083	0.0090	0.0090	0.1325	0.1316	0.1316
0.2	−0.0006	−0.0013	−0.0013	0.1350	0.1339	0.1339
0.4	0.0155	0.0128	0.0128	0.1349	0.1333	0.1333
0.6	0.0171	0.0124	0.0123	0.1447	0.1427	0.1427
0.8	0.0195	0.0126	0.0126	0.1551	0.1527	0.1527
1.0	0.0119	0.0029	0.0028	0.1602	0.1578	0.1578
1.2	0.0198	0.0080	0.0078	0.1801	0.1764	0.1763
1.4	0.0427	0.0267	0.0265	0.1972	0.1898	0.1895
1.6	0.0307	0.0097	0.0091	0.2136	0.2043	0.2038
1.8	0.0491	0.0187	0.0174	0.2475	0.2294	0.2280
2.0	0.0700	0.0249	0.0222	0.2889	0.2569	0.2538
2.2	0.0558	−0.0078	−0.0134	0.3458	0.2973	0.2895
2.4	0.0812	−0.0094	−0.0230	0.4217	0.3560	0.3305
2.6	0.1455	0.0154	−0.0133	0.5109	0.4145	0.3652
2.8	0.1251	−0.0305	−0.0758	0.5195	0.4355	0.3704
3.0	0.1763	−0.0120	−0.0977	0.5585	0.4737	0.3724

Table A12

Average Bias and RMSE of Ability Estimates With 200 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 500

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.1303	−0.0612	−0.0524	0.3643	0.3215	0.3050
−2.8	−0.1139	−0.0619	−0.0576	0.3250	0.2894	0.2813
−2.6	−0.1007	−0.0634	−0.0617	0.2708	0.2428	0.2411
−2.4	−0.0817	−0.0563	−0.0554	0.2381	0.2206	0.2197
−2.2	−0.0934	−0.0750	−0.0746	0.2142	0.2014	0.2009
−2.0	−0.0776	−0.0639	−0.0637	0.1951	0.1873	0.1871
−1.8	−0.0679	−0.0567	−0.0566	0.1779	0.1724	0.1723
−1.6	−0.0505	−0.0413	−0.0413	0.1586	0.1546	0.1546
−1.4	−0.0435	−0.0363	−0.0363	0.1453	0.1420	0.1420
−1.2	−0.0345	−0.0292	−0.0291	0.1306	0.1283	0.1283
−1.0	−0.0236	−0.0198	−0.0198	0.1165	0.1151	0.1151
−0.8	−0.0168	−0.0141	−0.0141	0.1138	0.1130	0.1130
−0.6	−0.0145	−0.0125	−0.0125	0.1096	0.1090	0.1090
−0.4	−0.0075	−0.0061	−0.0061	0.1028	0.1025	0.1025
−0.2	0.0031	0.0040	0.0040	0.1062	0.1059	0.1059
0.0	0.0077	0.0079	0.0079	0.1022	0.1019	0.1019
0.2	0.0107	0.0103	0.0103	0.1001	0.0997	0.0997
0.4	0.0207	0.0195	0.0195	0.1040	0.1034	0.1034
0.6	0.0245	0.0222	0.0222	0.1073	0.1063	0.1063
0.8	0.0362	0.0331	0.0331	0.1188	0.1174	0.1174
1.0	0.0318	0.0277	0.0277	0.1207	0.1191	0.1191
1.2	0.0503	0.0450	0.0450	0.1355	0.1328	0.1328
1.4	0.0602	0.0532	0.0532	0.1437	0.1398	0.1398
1.6	0.0636	0.0543	0.0542	0.1657	0.1604	0.1603
1.8	0.0839	0.0711	0.0709	0.1836	0.1753	0.1751
2.0	0.0971	0.0796	0.0793	0.2057	0.1936	0.1933
2.2	0.0974	0.0739	0.0734	0.2225	0.2067	0.2063
2.4	0.1164	0.0840	0.0831	0.2757	0.2518	0.2507
2.6	0.1479	0.1036	0.1017	0.3223	0.2876	0.2846
2.8	0.1783	0.1191	0.1148	0.3714	0.3254	0.3175
3.0	0.2242	0.1482	0.1359	0.4237	0.3677	0.3441

Table A13

Average Bias and RMSE of Ability Estimates With 50 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 1,000

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.0373	0.5297	4.0392	0.8082	0.6780	5.3750
−2.8	−0.1635	0.3779	3.5756	0.8269	0.5550	5.0078
−2.6	−0.1574	0.2880	2.7286	0.8351	0.5033	4.2311
−2.4	−0.2108	0.1706	2.1320	0.8363	0.4501	3.6621
−2.2	−0.1535	0.1128	1.3217	0.7585	0.4051	2.7676
−2.0	−0.1349	0.0544	0.8781	0.7115	0.3905	2.3126
−1.8	−0.1077	0.0032	0.3605	0.5839	0.3676	1.4270
−1.6	−0.1244	−0.0500	0.1839	0.5534	0.3911	1.2188
−1.4	−0.0850	−0.0490	0.0247	0.4386	0.3682	0.7096
−1.2	−0.0719	−0.0508	−0.0413	0.3549	0.3338	0.4008
−1.0	−0.0479	−0.0327	−0.0324	0.3245	0.3166	0.3159
−0.8	−0.0420	−0.0305	−0.0305	0.2880	0.2818	0.2820
−0.6	−0.0043	0.0035	0.0035	0.2612	0.2574	0.2575
−0.4	−0.0035	0.0017	0.0016	0.2445	0.2411	0.2412
−0.2	−0.0192	−0.0168	−0.0168	0.2286	0.2245	0.2246
0.0	−0.0102	−0.0118	−0.0119	0.2220	0.2175	0.2174
0.2	−0.0174	−0.0232	−0.0233	0.2165	0.2121	0.2121
0.4	−0.0251	−0.0356	−0.0357	0.2083	0.2044	0.2044
0.6	−0.0010	−0.0185	−0.0188	0.2305	0.2224	0.2221
0.8	−0.0237	−0.0485	−0.0495	0.2531	0.2437	0.2427
1.0	−0.0464	−0.0817	−0.0835	0.2690	0.2583	0.2573
1.2	−0.0377	−0.0920	−0.0981	0.3382	0.3052	0.2897
1.4	−0.0241	−0.1022	−0.1108	0.3844	0.3421	0.3273
1.6	−0.0060	−0.1167	−0.1401	0.4901	0.4143	0.3655
1.8	0.0504	−0.1064	−0.1581	0.6065	0.4844	0.3959
2.0	0.1054	−0.1045	−0.1991	0.6934	0.5279	0.4020
2.2	0.1507	−0.1144	−0.2698	0.7624	0.5665	0.4233
2.4	0.2342	−0.1065	−0.3494	0.8237	0.5852	0.4529
2.6	0.2521	−0.1480	−0.4782	0.8376	0.5951	0.5364
2.8	0.2809	−0.1831	−0.6208	0.8314	0.5944	0.6521
3.0	0.2576	−0.2519	−0.7957	0.8047	0.6080	0.8123

Table A14

Average Bias and RMSE of Ability Estimates With 100 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 1,000

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.0658	0.1500	0.4360	0.6366	0.5465	0.4874
−2.8	−0.0973	0.0889	0.2853	0.6108	0.4912	0.3714
−2.6	−0.1730	0.0002	0.1511	0.6433	0.4841	0.3138
−2.4	−0.1596	−0.0299	0.0541	0.5876	0.4404	0.3069
−2.2	−0.1289	−0.0434	−0.0036	0.5084	0.3908	0.3051
−2.0	−0.1034	−0.0558	−0.0420	0.4084	0.3438	0.3107
−1.8	−0.1044	−0.0744	−0.0692	0.3657	0.3255	0.3101
−1.6	−0.1044	−0.0842	−0.0830	0.3325	0.3059	0.3036
−1.4	−0.0928	−0.0795	−0.0793	0.2817	0.2730	0.2727
−1.2	−0.0789	−0.0684	−0.0684	0.2394	0.2334	0.2336
−1.0	−0.0689	−0.0609	−0.0609	0.2341	0.2294	0.2294
−0.8	−0.0687	−0.0627	−0.0627	0.2160	0.2123	0.2123
−0.6	−0.0499	−0.0457	−0.0457	0.1880	0.1858	0.1858
−0.4	−0.0550	−0.0521	−0.0521	0.1828	0.1808	0.1808
−0.2	−0.0642	−0.0628	−0.0628	0.1736	0.1718	0.1718
0.0	−0.0621	−0.0628	−0.0628	0.1687	0.1674	0.1674
0.2	−0.0725	−0.0755	−0.0756	0.1650	0.1647	0.1647
0.4	−0.0819	−0.0873	−0.0873	0.1741	0.1750	0.1751
0.6	−0.0794	−0.0874	−0.0875	0.1824	0.1841	0.1841
0.8	−0.0872	−0.0980	−0.0980	0.1887	0.1916	0.1916
1.0	−0.1055	−0.1195	−0.1196	0.2108	0.2150	0.2150
1.2	−0.1084	−0.1271	−0.1272	0.2251	0.2301	0.2302
1.4	−0.1065	−0.1305	−0.1306	0.2349	0.2413	0.2413
1.6	−0.1079	−0.1394	−0.1404	0.2737	0.2752	0.2744
1.8	−0.1127	−0.1561	−0.1588	0.3087	0.3047	0.3033
2.0	−0.1054	−0.1682	−0.1756	0.3617	0.3464	0.3362
2.2	−0.1161	−0.2057	−0.2173	0.4285	0.3991	0.3888
2.4	−0.0527	−0.1833	−0.2140	0.5329	0.4806	0.4356
2.6	−0.0703	−0.2311	−0.2766	0.5638	0.5271	0.4749
2.8	0.0004	−0.2110	−0.2908	0.6295	0.5771	0.5032
3.0	0.0127	−0.2224	−0.3517	0.6417	0.6231	0.5336

Table A15

Average Bias and RMSE of Ability Estimates With 200 3PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 1,000

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.1360	0.0032	0.1091	0.5360	0.4520	0.3301
−2.8	−0.1649	−0.0436	0.0169	0.5201	0.4120	0.3185
−2.6	−0.1861	−0.0929	−0.0562	0.5029	0.4045	0.3329
−2.4	−0.1784	−0.1095	−0.0959	0.4533	0.3643	0.3343
−2.2	−0.1518	−0.1126	−0.1075	0.3632	0.3172	0.3033
−2.0	−0.1486	−0.1237	−0.1227	0.3166	0.2867	0.2856
−1.8	−0.1424	−0.1259	−0.1255	0.2925	0.2736	0.2726
−1.6	−0.1393	−0.1279	−0.1279	0.2582	0.2490	0.2489
−1.4	−0.1288	−0.1202	−0.1202	0.2272	0.2205	0.2206
−1.2	−0.1171	−0.1109	−0.1110	0.2033	0.1981	0.1982
−1.0	−0.1119	−0.1081	−0.1081	0.1916	0.1881	0.1881
−0.8	−0.1145	−0.1124	−0.1124	0.1853	0.1831	0.1832
−0.6	−0.1095	−0.1085	−0.1085	0.1738	0.1727	0.1727
−0.4	−0.1129	−0.1124	−0.1124	0.1683	0.1677	0.1677
−0.2	−0.1245	−0.1245	−0.1245	0.1737	0.1735	0.1735
0.0	−0.1255	−0.1262	−0.1262	0.1713	0.1716	0.1716
0.2	−0.1396	−0.1413	−0.1413	0.1778	0.1787	0.1787
0.4	−0.1471	−0.1498	−0.1498	0.1878	0.1895	0.1896
0.6	−0.1482	−0.1521	−0.1521	0.1913	0.1938	0.1938
0.8	−0.1564	−0.1616	−0.1616	0.1969	0.2005	0.2005
1.0	−0.1661	−0.1728	−0.1728	0.2132	0.2178	0.2177
1.2	−0.1701	−0.1786	−0.1786	0.2185	0.2243	0.2243
1.4	−0.1773	−0.1879	−0.1879	0.2298	0.2370	0.2370
1.6	−0.1753	−0.1886	−0.1885	0.2431	0.2511	0.2510
1.8	−0.1727	−0.1901	−0.1903	0.2555	0.2640	0.2640
2.0	−0.1845	−0.2079	−0.2084	0.2771	0.2871	0.2872
2.2	−0.1940	−0.2261	−0.2270	0.3039	0.3161	0.3164
2.4	−0.1853	−0.2305	−0.2318	0.3408	0.3511	0.3513
2.6	−0.1778	−0.2389	−0.2413	0.3875	0.3944	0.3928
2.8	−0.1363	−0.2188	−0.2238	0.4147	0.4157	0.4108
3.0	−0.1173	−0.2188	−0.2330	0.4515	0.4552	0.4372

Table A16

Average Bias and RMSE of Ability Estimates With 50 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 1,000

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.3501	−0.0673	0.2693	0.6615	0.4552	0.3244
−2.8	−0.3482	−0.1093	0.1367	0.6912	0.4899	0.2717
−2.6	−0.3504	−0.1483	0.0135	0.6862	0.4918	0.2772
−2.4	−0.2604	−0.1130	−0.0317	0.6115	0.4639	0.3261
−2.2	−0.2708	−0.1530	−0.1110	0.5712	0.4467	0.3632
−2.0	−0.2123	−0.1261	−0.1031	0.5141	0.4276	0.3695
−1.8	−0.1385	−0.0776	−0.0723	0.4051	0.3593	0.3458
−1.6	−0.0711	−0.0270	−0.0256	0.3169	0.2930	0.2920
−1.4	−0.0811	−0.0471	−0.0460	0.2958	0.2740	0.2735
−1.2	−0.0459	−0.0232	−0.0224	0.2482	0.2343	0.2338
−1.0	−0.0278	−0.0134	−0.0130	0.2141	0.2059	0.2056
−0.8	−0.0242	−0.0148	−0.0146	0.2105	0.2053	0.2051
−0.6	−0.0122	−0.0060	−0.0059	0.2021	0.1994	0.1994
−0.4	−0.0141	−0.0097	−0.0097	0.1936	0.1919	0.1919
−0.2	−0.0140	−0.0111	−0.0111	0.1882	0.1864	0.1864
0.0	−0.0025	−0.0019	−0.0019	0.1933	0.1910	0.1909
0.2	−0.0050	−0.0073	−0.0073	0.1875	0.1845	0.1845
0.4	0.0115	0.0052	0.0051	0.1873	0.1831	0.1830
0.6	0.0105	−0.0004	−0.0007	0.2052	0.1998	0.1996
0.8	0.0199	0.0034	0.0030	0.2257	0.2183	0.2181
1.0	−0.0025	−0.0244	−0.0249	0.2292	0.2231	0.2230
1.2	0.0274	−0.0025	−0.0032	0.2628	0.2514	0.2509
1.4	0.0649	0.0246	0.0230	0.3112	0.2884	0.2860
1.6	0.0580	0.0057	0.0022	0.3504	0.3159	0.3093
1.8	0.0761	0.0061	−0.0053	0.4020	0.3491	0.3207
2.0	0.1497	0.0420	0.0120	0.5076	0.4046	0.3411
2.2	0.1451	0.0050	−0.0504	0.5582	0.4361	0.3365
2.4	0.1873	−0.0004	−0.1076	0.6449	0.4959	0.3447
2.6	0.2787	0.0379	−0.1593	0.7101	0.5377	0.3253
2.8	0.2590	−0.0195	−0.2781	0.6942	0.5365	0.3741
3.0	0.2757	−0.0383	−0.4037	0.6862	0.5421	0.4548

Table A17

Average Bias and RMSE of Ability Estimates With 100 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 1,000

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.2675	−0.1277	−0.0622	0.5281	0.4363	0.3347
−2.8	−0.2364	−0.1205	−0.0878	0.5038	0.4087	0.3510
−2.6	−0.2247	−0.1349	−0.1196	0.4642	0.3733	0.3448
−2.4	−0.1660	−0.1075	−0.1016	0.3759	0.3204	0.3100
−2.2	−0.1724	−0.1304	−0.1271	0.3486	0.3080	0.3003
−2.0	−0.1372	−0.1082	−0.1072	0.2933	0.2706	0.2691
−1.8	−0.1170	−0.0953	−0.0948	0.2592	0.2445	0.2439
−1.6	−0.0893	−0.0725	−0.0723	0.2276	0.2178	0.2177
−1.4	−0.0821	−0.0689	−0.0688	0.2091	0.2014	0.2013
−1.2	−0.0669	−0.0569	−0.0568	0.1940	0.1884	0.1884
−1.0	−0.0460	−0.0386	−0.0386	0.1696	0.1661	0.1661
−0.8	−0.0371	−0.0315	−0.0315	0.1617	0.1594	0.1593
−0.6	−0.0337	−0.0294	−0.0294	0.1547	0.1530	0.1530
−0.4	−0.0229	−0.0196	−0.0196	0.1466	0.1455	0.1455
−0.2	−0.0125	−0.0103	−0.0103	0.1409	0.1399	0.1399
0.0	−0.0037	−0.0029	−0.0029	0.1345	0.1335	0.1335
0.2	−0.0083	−0.0091	−0.0091	0.1372	0.1361	0.1361
0.4	0.0121	0.0094	0.0094	0.1365	0.1350	0.1350
0.6	0.0173	0.0126	0.0125	0.1464	0.1444	0.1444
0.8	0.0220	0.0150	0.0149	0.1565	0.1539	0.1539
1.0	0.0179	0.0084	0.0083	0.1633	0.1602	0.1602
1.2	0.0306	0.0177	0.0175	0.1848	0.1797	0.1795
1.4	0.0601	0.0421	0.0417	0.2073	0.1974	0.1972
1.6	0.0553	0.0314	0.0308	0.2286	0.2161	0.2156
1.8	0.0892	0.0552	0.0539	0.2730	0.2499	0.2485
2.0	0.1223	0.0736	0.0709	0.3232	0.2840	0.2806
2.2	0.1210	0.0540	0.0483	0.3790	0.3215	0.3116
2.4	0.1607	0.0679	0.0531	0.4574	0.3822	0.3512
2.6	0.2370	0.1074	0.0752	0.5492	0.4442	0.3841
2.8	0.2205	0.0664	0.0174	0.5507	0.4492	0.3711
3.0	0.2748	0.0956	0.0023	0.5807	0.4824	0.3575

Table A18

Average Bias and RMSE of Ability Estimates With 200 2PL Items and 1,000 Students at Each Ability Level When Item Parameters Are Estimated With Calibration Sample Size 1,000

	Average bias			RMSE		
Ability	MLE	WLE	MLE-LBC	MLE	WLE	MLE-LBC
−3.0	−0.1905	−0.1227	0.1131	0.3897	0.3404	0.3218
−2.8	−0.1724	−0.1205	0.1161	0.3560	0.3148	0.3058
−2.6	−0.1504	−0.1128	0.1112	0.2980	0.2664	0.2643
−2.4	−0.1248	−0.0981	0.0973	0.2625	0.2422	0.2413
−2.2	−0.1281	−0.1081	0.1077	0.2370	0.2218	0.2214
−2.0	−0.1071	−0.0919	0.0917	0.2106	0.2003	0.2001
−1.8	−0.0941	−0.0821	0.0820	0.1900	0.1824	0.1823
−1.6	−0.0752	−0.0658	0.0658	0.1685	0.1630	0.1630
−1.4	−0.0669	−0.0596	0.0596	0.1544	0.1502	0.1501
−1.2	−0.0566	−0.0512	0.0512	0.1390	0.1359	0.1359
−1.0	−0.0445	−0.0407	0.0406	0.1233	0.1213	0.1213
−0.8	−0.0362	−0.0334	0.0334	0.1191	0.1178	0.1178
−0.6	−0.0314	−0.0294	0.0294	0.1133	0.1124	0.1124
−0.4	−0.0227	−0.0213	0.0213	0.1060	0.1055	0.1055
−0.2	−0.0110	−0.0101	0.0101	0.1074	0.1070	0.1070
0.0	−0.0047	−0.0045	0.0045	0.1025	0.1022	0.1022
0.2	−0.0003	−0.0007	0.0007	0.0997	0.0993	0.0993
0.4	0.0110	0.0097	0.0097	0.1025	0.1020	0.1019
0.6	0.0151	0.0129	0.0129	0.1055	0.1047	0.1047
0.8	0.0269	0.0236	0.0236	0.1157	0.1145	0.1145
1.0	0.0232	0.0190	0.0190	0.1192	0.1179	0.1179
1.2	0.0424	0.0369	0.0368	0.1324	0.1298	0.1298
1.4	0.0541	0.0468	0.0468	0.1421	0.1382	0.1382
1.6	0.0600	0.0503	0.0502	0.1663	0.1610	0.1610
1.8	0.0862	0.0729	0.0728	0.1878	0.1793	0.1792
2.0	0.1021	0.0844	0.0841	0.2106	0.1985	0.1983
2.2	0.1053	0.0821	0.0817	0.2286	0.2129	0.2124
2.4	0.1284	0.0966	0.0957	0.2820	0.2574	0.2563
2.6	0.1588	0.1152	0.1133	0.3249	0.2886	0.2858
2.8	0.1881	0.1295	0.1247	0.3724	0.3249	0.3159
3.0	0.2327	0.1559	0.1431	0.4246	0.3651	0.3410