Investigating the Impact of Uncertainty About Item Parameters on Ability Estimation

Jinming Zhang

University of Illinois at Urbana-Champaign

Minge Xie

Rutgers University

Xiaolan Song

JP Morgan Chase

Ting Lu

University of Illinois at Urbana-Champaign

The research was partially supported by Grants NSF SES-0241859 and ONR N00014-05-1-0237. Requests for reprints should be sent to Jinming Zhang, 236A Education Building, Department of Educational Psychology, 1310 S Sixth Street, Champaign, IL 61820 (E-mail: jmzhang@illinois.edu).

Investigating the Impact of Uncertainty About Item Parameters on Ability Estimation

Abstract

Asymptotic expansions of the maximum likelihood estimator (MLE) and weighted likelihood estimator (WLE) of an examinee's ability are derived while item parameter estimators are treated as covariates measured with error. The asymptotic formulae present the amount of bias of the ability estimators due to the uncertainty of item parameter estimators. A numerical example is presented to illustrate how to apply the formulae to evaluate the impact of uncertainty about item parameters on ability estimation and the appropriateness of estimating ability using the regular MLE or WLE method.

Key words: bias, item response theory (IRT), measurement error, maximum likelihood estimator (MLE), weighted likelihood estimator (WLE).

1. Introduction

In practical applications of item response theory (IRT), item parameters are usually estimated first from a calibration sample. After treating these estimates as true values of item parameters, ability parameters are then estimated and further statistical inferences are made. When item parameter estimation is sufficiently accurate, it may not be problematic to substitute the estimated item parameters for the true ones in the IRT models in estimating ability parameter. However, when the measurement errors in item parameter estimators are no longer ignorable, the statistical inferences based on such a substitution could be misleading. For instance, Tsutakawa and Johnson (1990) demonstrated that both the maximum likelihood and empirical Bayes approaches underestimate the variance of ability when the uncertainty of item parameter estimation is ignored. This is true especially when calibration samples are small.

Various approaches have been proposed to address the bias in ability estimation resulting from the uncertainty of item parameter estimation (Lewis, 1985, 2001; Mislevy, Wingersky, & Sheehan, 1994; Song, 2003; Tsutakawa & Johnson, 1990). One of these approaches, the measurement error approach, treats item parameter estimators as covariates measured with errors instead of treating them as being fixed, so the impact of uncertainty about item parameters can be investigated along the line of what has been done in research on measurement error models (Stefanski & Carroll, 1985). The study of measurement error models first appeared in the late 19th century, and derived from the fact that quantitative variables are seldom measured with complete precision. Reviews and discussions of measurement error models were given by Fuller (1987) and Carroll, Ruppert, Stefanski, and Crainiceanu (2006).

In this paper, linear models are used to characterize the measurement errors in item parameter estimators. Measurement errors include chance error (or random error) and systematic error (or bias). Under certain regularity conditions, we derive asymptotic expansion formulae for the maximum likelihood estimator (MLE) and weighted likelihood estimator (WLE) of ability when measurement errors in item parameter estimators are presented. These expansion formulae present the major part of the bias (called the bias function in this paper) in the MLE and WLE of ability, when item parameters are estimated with measurement errors. The bias function of the MLE and WLE of ability is a function of ability with the bias, variance, and covariance of item parameter estimators as its parameters. A numerical example is presented to demonstrate how to use the bias function to examine the impact of uncertainty about item parameters on the MLE and WLE of ability. The degree of uncertainty about item parameters in this paper refers to the amount of bias, variance, and covariance (or measurement errors) of item parameter estimators.

2. Ability Estimation When Item Parameters Are Known

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) for $i = 1, \ldots, n$. The item response function (IRF) of a three-parameter logistic (3PL) model is

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

where θ is the ability parameter, a_i, b_i, and c_i are the item discrimination, difficulty, and guessing parameters, respectively. This 3PL model will be used throughout this paper unless specified otherwise. The following notations are introduced for the simplicity of expression for the formulae presented in this paper. The reader may refer back here later.

$$\begin{aligned}
Q_i(\theta) &= 1 - P_i(\theta), & (2)\\
P_i^*(\theta) &= \frac{1}{1 + \exp\{-1.7a_i(\theta - b_i)\}}, & (3)\\
Q_i^*(\theta) &= 1 - P_i^*(\theta), & (4)\\
K_i(\theta) &= \frac{P_i^*(\theta)}{P_i(\theta)} = \frac{1}{1 + c_i \exp\{-1.7a_i(\theta - b_i)\}}, & (5)\\
L_i(\theta) &= \frac{Q_i^*(\theta)}{P_i(\theta)} = \frac{1}{c_i + \exp\{1.7a_i(\theta - b_i)\}}, & (6)\\
M_i(\theta) &= \frac{1}{2} - P_i^*(\theta) + c_i L_i(\theta), & (7)\\
Z_n(\theta; \mathbf{y}) &= 1.7 \sum_{i=1}^{n} a_i K_i(\theta)(y_i - P_i(\theta)). & (8)
\end{aligned}$$

Note that $P_i^*(\theta)$ is a two-parameter logistic (2PL) model, which is mainly used as a notation in this paper. Clearly, $P_i(\theta) = c_i + (1 - c_i)P_i^*(\theta)$, and the first derivative of $P_i(\theta)$ with respect to θ can be expressed as

$$P_i^{'}(\theta) = 1.7a_i P_i^*(\theta) Q_i(\theta). \qquad (9)$$

Under the assumptions of *local* or *conditional* independence (Lord, 1980), the likelihood function for the response vector $\mathbf{y}$ is

$$L(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} P_i^{y_i}(\theta)\, Q_i^{1-y_i}(\theta). \tag{10}$$

Assume that item parameters (a_i, b_i, c_i) in these models are known and fixed, the MLE $\hat{\theta}_m$ is defined as the value of θ that maximizes (10). In practice, $\hat{\theta}_m$ is often found by setting the derivative of the likelihood function to zero; that is, $\hat{\theta}_m$ satisfies

$$\frac{\partial \ln L(\mathbf{y} \mid \theta)}{\partial \theta} = \sum_{i=1}^{n} \left(\frac{y_i - P_i(\theta)}{P_i(\theta) Q_i(\theta)} \right) P_i'(\theta) = 0. \tag{11}$$

By (5) and (9), the likelihood equation (11) becomes

$$Z_n(\theta; \mathbf{y}) \equiv 1.7 \sum_{i=1}^{n} a_i K_i(\theta)(y_i - P_i(\theta)) = 0. \tag{12}$$

Let

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

be the Fisher test information function. The variance of the MLE $\hat{\theta}_m$ is $\text{Var}(\hat{\theta}_m) = 1/I_n(\hat{\theta}_m)$. When all the items are modeled by 3PL models,

$$I_n(\theta) = 1.7^2 \sum_{i=1}^{n} a_i^2 (1 - c_i) P_i^*(\theta) Q_i^*(\theta) K_i(\theta). \tag{13}$$

The i-*th* term in (13) is the item information function of item i,

$$I_{(i)}(\theta) = 1.7^2 a_i^2 (1 - c_i) P_i^*(\theta) Q_i^*(\theta) K_i(\theta), \tag{14}$$

and

$$I_n(\theta) = \sum_{i=1}^{n} I_{(i)}(\theta).$$

Given item parameters in 3PL models as known, Lord (1983) obtained the following bias function for the MLE of θ:

$$B_n(\theta) = \frac{1.7}{I_n^2(\theta)} \sum_{i=1}^{n} a_i I_{(i)}(\theta) \left(P_i^*(\theta) - \frac{1}{2} \right). \tag{15}$$

The MLE with Lord bias-correction of θ is defined as

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

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

Based on Lord's work, Warm (1989) proposed the weighted likelihood estimation method and showed that the WLE of ability is less biased than the MLE with the same asymptotic variance and normal distribution under the same assumption that the true item parameters are known. The WLE $\hat{\theta}_w$ is defined as the value of θ that maximizes

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

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

$$\frac{\partial \ln f(\theta)}{\partial \theta} = -B_n(\theta)I_n(\theta).$$

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

$$\frac{\partial \ln[f(\theta)L(\mathbf{y} \mid \theta)]}{\partial \theta} = \sum_{i=1}^{n} \left(\frac{y_i - P_i(\theta)}{P_i(\theta)Q_i(\theta)} \right) P_i^{'}(\theta) - B_n(\theta)I_n(\theta) = 0.$$

That is,

$$Z_n(\theta; \mathbf{y}) - B_n(\theta)I_n(\theta) = 0. \tag{16}$$

Warm (1989) proved that the bias of $\hat{\theta}_w$, BIAS($\hat{\theta}_w$), is $o(n^{-1})$ under the assumption that the item parameters are known.

3. Ability Estimation When Item Parameters Are Estimated

In reality, no true item parameters or ability parameters are known. As mentioned in the first section, it is a common practice to estimate item parameters first and then to estimate ability parameters, treating those previously estimated item parameters as if they were the true quantities. That is, the MLE of an ability parameter is obtained by assuming item parameter estimators $\hat{a}_i$, $\hat{b}_i$, and $\hat{c}_i$ are the true item parameters. Hence, the MLE $\hat{\theta}_m$ satisfies

$$\hat{Z}_n(\hat{\theta}_m; \mathbf{y}) = 1.7\sum_{i=1}^{n} \hat{a}_i \hat{K}_i(\hat{\theta}_m)(y_i - \hat{P}_i(\hat{\theta}_m)) = 0 \tag{17}$$

rather than (12), where $\hat{P}_i(\theta) = P(\theta; \hat{a}_i, \hat{b}_i, \hat{c}_i)$ and $\hat{K}_i(\theta) = K(\theta; \hat{a}_i, \hat{b}_i, \hat{c}_i)$, while the WLE $\hat{\theta}_w$ satisfies

$$\hat{Z}_n(\hat{\theta}_w; \mathbf{y}) - \hat{B}_n(\hat{\theta}_w)\hat{I}_n(\hat{\theta}_w) = 0 \tag{18}$$

rather than (16). In this paper, $\hat{\theta}_m$ satisfying (17) and $\hat{\theta}_w$ satisfying (18) (based on fixed estimated item parameters) are called the naive MLE and the naive WLE, respectively. In practice, the naive MLE or the naive WLE is used because the (true) MLE satisfying (12) or the (true) WLE satisfying (16) based on the true values of item parameters cannot be obtained.

Suppose that item parameters are estimated using a calibration sample with N examinees. The item parameter estimators, $\hat{a}_i$, $\hat{b}_i$ and $\hat{c}_i$, are related to N. The label N is usually suppressed in these and other related quantities for convenience, unless necessary. These estimators are bound to variances and covariances. In addition, some amount of bias is almost inevitable. Accordingly, this paper treats item parameter estimators as covariates measured with error in ability estimation. The measurement error models used in this paper are

$$\begin{aligned} \hat{a}_i &= a_i + \delta_{ai} + \varepsilon_{ai}, &(19)\\ \hat{b}_i &= b_i + \delta_{bi} + \varepsilon_{bi}, &(20)\\ \hat{c}_i &= c_i + \delta_{ci} + \varepsilon_{ci}, &(21) \end{aligned}$$

where δ_{ai}, δ_{bi}, and δ_{ci} are the biases of corresponding item parameter estimators, and $\{(\varepsilon_{ai}, \varepsilon_{bi}, \varepsilon_{ci})\}$ is an independent sequence of random error vectors[1] with mean zero and covariance matrix

$$\Sigma_i = \begin{pmatrix} \sigma_{ai}^2 & \sigma_{abi} & \sigma_{aci} \\ \sigma_{abi} & \sigma_{bi}^2 & \sigma_{bci} \\ \sigma_{aci} & \sigma_{bci} & \sigma_{ci}^2 \end{pmatrix}.$$

That is, σ_{ai}, σ_{bi} and σ_{ci} are the standard errors of item parameter estimators while σ_{abi}, σ_{bci} and σ_{aci} are the corresponding covariances. There is no requirement for the distribution of $(\varepsilon_{ai}, \varepsilon_{bi}, \varepsilon_{bi})$ in these models considered in this paper. When item parameter estimators are unbiased, all of the δs are zero (i.e., $\delta_{ai} = \delta_{bi} = \delta_{ci} = 0$). When estimated item parameters are treated as the true values (i.e., $\hat{a}_i = a_i$, $\hat{b}_i = b_i$, and $\hat{c}_i = c_i$), like the cases considered in Lord (1983) and Warm (1989), all of the δs and σs (or εs) are zero.

[1]In Bayesian setting, item parameters are usually assumed to be independent between items, that is, $\{(a_i, b_i, c_i)\}$ is an independent sequence of random vectors (see Lewis, 2001). Lewis argued that this is almost a necessary condition. In practice, only the covariances of item parameter estimators within an item are available and the covariances of item parameter estimators between items are zero. Thus, it is not too unreasonable to assume that $\{(\varepsilon_{ai}, \varepsilon_{bi}, \varepsilon_{bi})\}$ is an independent sequence of random vectors.

The theoretical results in this paper require the following *regularity* conditions. These conditions and their explanations or justifications are presented first.

Regularity Conditions

(C0) Item parameters a_i and b_i are uniformly bounded and c_i is bounded away from 1. θ is a bounded variable.

(C1) There exists n_0 such that for any $n > n_0$,

$$\lim_{N\to\infty} \sigma_n^2 = 0,$$

where

$$\sigma_n^2 \overset{\text{def}}{=} \sigma_{nN}^2 = \max_{1\le i\le n}\{\sigma_{ai}^2, \sigma_{bi}^2, \sigma_{ci}^2, \delta_{ai}^2, \delta_{bi}^2, \delta_{ci}^2\}.$$

Recall that n is the number of items and N is the calibration sample size. σ_{nN}^2 is related to N, because it is the maximum value of the variances and squared biases of item parameter estimators from the calibration sample. As a convention in this paper, the subscript N is suppressed from σ_{nN}^2.

(C2)

$$\lim_{N\to\infty}\frac{1}{n}\sum_{i=1}^{n}\text{Var}[(\hat{a}_i - a_i)^2] = 0, \quad \lim_{N\to\infty}\frac{1}{n}\sum_{i=1}^{n}\text{Var}[(\hat{b}_i - b_i)^2] = 0,$$

$$\lim_{N\to\infty}\frac{1}{n}\sum_{i=1}^{n}\text{Var}[(\hat{a}_i - a_i)(\hat{b}_i - b_i)] = 0, \quad \lim_{N\to\infty}\frac{1}{n}\sum_{i=1}^{n}\text{Var}[(\hat{c}_i - c_i)^2] = 0,$$

$$\lim_{N\to\infty}\frac{1}{n}\sum_{i=1}^{n}\text{Var}[(\hat{a}_i - a_i)(\hat{c}_i - c_i)] = 0, \quad \lim_{N\to\infty}\frac{1}{n}\sum_{i=1}^{n}\text{Var}[(\hat{b}_i - b_i)(\hat{c}_i - c_i)] = 0.$$

(C3) $(\hat{a}_i - a_i)/\sigma_{ai}$, $(\hat{b}_i - b_i)/\sigma_{bi}$, and $(\hat{c}_i - c_i)/\sigma_{ci}$ have uniformly bounded 4 moments.

(C4) For any fixed θ, there exists $c_0(\theta) > 0$ such that

$$\liminf_{n\to\infty} I_n(\theta)/n \ge c_0(\theta) > 0.$$

In effect, (C0), which is also required by Lord (1983), holds in all applications. Regularity Condition (C1) states that the biases and standard errors of item parameter estimators converge to zero when the calibration sample size tends to infinity, which means that item parameter estimators are asymptotically consistent. (C2) also requires that item parameter estimators be reasonably accurate when the calibration sample size is large enough. Regularity Condition (C3) is a very weak assumption under (C0). (C4) requires the values of the test information function to be reasonably large, which should hold for all well-designed tests when θ is bounded. In fact, this condition is commonly assumed. For example, Chang and Stout (1993) also required the same condition in proving the asymptotic posterior normality of the latent ability.

Note that Regularity Conditions (C1), (C2) and (C3) are much weaker than the condition that item parameters are known and fixed (i.e., all of the δs and σs are zero) as assumed by Lord (1983) and Warm (1989).

Under the regularity conditions, we obtain the following asymptotic expansion result for the MLE and WLE of ability. In the theorem, notations $o_p(\cdot)$ and $O_p(\cdot)$ are needed, so that $F_n = G_n + o_p(H_n)$ means that $(F_n - G_n)/H_n$ converges to zero in probability, and $F_n = O_p(1)$ means that $\{F_n\}$ are bounded in probability, whereas $o(\cdot)$ and $O(\cdot)$ are in regular sense (see Serfling, 1980).

Theorem. Suppose that $\hat{\theta}_m$ is the naive MLE of θ and satisfies (17) and $\hat{\theta}_w$ is the naive WLE of θ and satisfies (18). Assume that Regularity Conditions (C0)–(C4) hold. Then

$$\hat{\theta}_m = \theta + [D_n(\theta) + J_n(\theta) + Z_n(\theta; \mathbf{y})]/I_n(\theta) + o_p\left(\max\left(\sigma_n^2, \frac{1}{\sqrt{n}}\right)\right), \tag{22}$$

and

$$\hat{\theta}_w = \theta + [D_n(\theta) + J_n(\theta) + Z_n(\theta; \mathbf{y}) - B_n(\theta)I_n(\theta)]/I_n(\theta) + o_p\left(\max\left(\sigma_n^2, \frac{1}{\sqrt{n}}\right)\right), \tag{23}$$

where $I_n(\theta)$ is the Fisher test information function given by (13), $B_n(\theta)$ is given by (15),

$Z_n(\theta; \mathbf{y})$ is given by (8),

$$
\begin{aligned}
D_n(\theta) &= D_{n,1}(\theta) + D_{n,2}(\theta) + D_{n,3}(\theta), \\
D_{n,1}(\theta) &= -\sum_{i=1}^{n} \frac{(\theta - b_i)}{a_i} I_{(i)}(\theta) \delta_{ai}, \\
D_{n,2}(\theta) &= \sum_{i=1}^{n} I_{(i)}(\theta) \delta_{bi}, \\
D_{n,3}(\theta) &= -1.7 \sum_{i=1, c_i>0}^{n} a_i Q_i^*(\theta) K_i(\theta) \delta_{ci}, \\
J_n(\theta) &= J_{n,1}(\theta) + J_{n,2}(\theta) + J_{n,3}(\theta) + J_{n,4}(\theta) + J_{n,5}(\theta) + J_{n,6}(\theta), \\
J_{n,1}(\theta) &= -\sum_{i=1}^{n} \frac{(\theta - b_i)}{a_i^2} I_{(i)}(\theta) \left\{1.7 a_i (\theta - b_i) M_i(\theta) + 1\right\} (\sigma_{ai}^2 + \delta_{ai}^2), \\
J_{n,2}(\theta) &= -1.7 \sum_{i=1}^{n} a_i I_{(i)}(\theta) M_i(\theta) (\sigma_{bi}^2 + \delta_{bi}^2), \\
J_{n,3}(\theta) &= \sum_{i=1}^{n} \frac{2}{a_i} I_{(i)}(\theta) \left\{1.7 a_i (\theta - b_i) M_i(\theta) + 1\right\} (\sigma_{abi} + \delta_{ai}\delta_{bi}), \\
J_{n,4}(\theta) &= 1.7 \sum_{i=1,\ c_i>0}^{n} a_i Q_i^*(\theta) K_i(\theta) L_i(\theta) (\sigma_{ci}^2 + \delta_{ci}^2), \\
J_{n,5}(\theta) &= 1.7 \sum_{i=1\ c_i>0}^{n} Q_i^*(\theta) K_i(\theta) \left\{1.7 a_i (\theta - b_i)[1 - 2c_i L_i(\theta)] - 1\right\} (\sigma_{aci} + \delta_{ai}\delta_{ci}), \\
J_{n,6}(\theta) &= -1.7^2 \sum_{i=1\ c_i>0}^{n} a_i^2 Q_i^*(\theta) K_i(\theta) [1 - 2c_i L_i(\theta)] (\sigma_{bci} + \delta_{bi}\delta_{ci}),
\end{aligned}
$$

and $I_{(i)}(\theta)$ is the item information function of item i.

The proof of this theorem is given in the appendix. The theorem provides the bias of the naive MLE and WLE of ability obtained by treating estimated item parameters as though they were the true values. The formulae are rather complicated, but they are quite general to include one-parameter logistic (1PL) and 2PL models as special cases. The 1PL model is

$$P_i^{**}(\theta) = \frac{1}{1 + \exp\{-1.7(\theta - b_i)\}}, \tag{24}$$

when $a_i = 1$ in (3). If item i is modeled by 2PL, then all the quantities related to c_i are zero (e.g., $\hat{c}_i = c_i = \delta_{ci} = \varepsilon_{ci} = \sigma_{aci} = \sigma_{bci} = \sigma_{ci}^2 = 0$). If item i is modeled by 1PL, then $\hat{b}_i = b_i = 1$ and all other quantities related to b_i and/or c_i are zero. Thus, the theorem can be applied to cases where different models are used to characterize different types of items in a test. The formulae in the theorem can be much more simplified if all items in a test are modeled by 1PL or 2PL models. The results for these two special cases are presented below.

Corollary 1. Under the same conditions of the Theorem, if all items are modeled by 1PL models, then (22) and (23) hold. That is,

$$\hat{\theta}_m = \theta + [D_n(\theta) + J_n(\theta) + Z_n(\theta; \mathbf{y})]/I_n(\theta) + o_p\left(\max\left(\sigma_n^2, \frac{1}{\sqrt{n}}\right)\right),$$

and

$$\hat{\theta}_w = \theta + [D_n(\theta) + J_n(\theta) + Z_n(\theta; \mathbf{y}) - B_n(\theta)I_n(\theta)]/I_n(\theta) + o_p\left(\max\left(\sigma_n^2, \frac{1}{\sqrt{n}}\right)\right),$$

where

$$D_n(\theta) = \sum_{i=1}^{n} I_{(i)}(\theta)\delta_{bi}, \tag{25}$$

$$J_n(\theta) = 1.7\sum_{i=1}^{n}\left[P_i^{**}(\theta) - \frac{1}{2}\right] I_{(i)}(\theta)(\sigma_{bi}^2 + \delta_{bi}^2), \tag{26}$$

$I_{(i)}(\theta)$ is the item information function for the 1PL model, $I_n(\theta) = \sum_{i=1}^{n} I_{(i)}(\theta)$ is the test information function, and $Z_n(\theta; \mathbf{y}) = 1.7\sum_{i=1}^{n}(y_i - P_i^{**}(\theta))$.

Note that in the 1PL case, $D_n(\theta)$ is a linear combination of biases of item difficulty parameter estimators with item information function as its coefficients, and $J_n(\theta)$ is a linear combination of variances and squared biases of item difficulty parameter estimators (i.e., the mean squared errors (MSEs) of $\hat{b}_i$). Notice that $P_i^{**}(\theta) > 0.5$ when $\theta > b_i$ and $P_i^{**}(\theta) < 0.5$ when $\theta < b_i$. Thus, when θ is above the range of item difficulty parameters, all the coefficients in $J_n(\theta)$ will be positive, and therefore $J_n(\theta)$ can be relatively large. Contrarily, when θ is below the range of item difficulty parameters, $J_n(\theta)$ will be negative with a relatively large (absolute) value.

Corollary 2. Under the same conditions of the Theorem, if all items are modeled by 2PL models, then (22) and (23) hold. Here

$$\begin{aligned}
D_n(\theta) &= \sum_{i=1}^{n} I_{(i)}(\theta)\delta_{bi} - \sum_{i=1}^{n} \frac{(\theta - b_i)}{a_i} I_{(i)}(\theta)\delta_{ai}, \\
J_n(\theta) &= J_{n,1}(\theta) + J_{n,2}(\theta) + J_{n,3}(\theta), \\
J_{n,1}(\theta) &= -\sum_{i=1}^{n} \frac{(\theta - b_i)}{a_i^2} \left\{1 - 1.7a_i(\theta - b_i)\left[P_i^*(\theta) - \frac{1}{2}\right]\right\} I_{(i)}(\theta)(\sigma_{ai}^2 + \delta_{ai}^2), \\
J_{n,2}(\theta) &= 1.7\sum_{i=1}^{n} a_i \left[P_i^*(\theta) - \frac{1}{2}\right] I_{(i)}(\theta)(\sigma_{bi}^2 + \delta_{bi}^2), \\
J_{n,3}(\theta) &= \sum_{i=1}^{n} \frac{2}{a_i} \left\{1 - 1.7a_i(\theta - b_i)\left[P_i^*(\theta) - \frac{1}{2}\right]\right\} I_{(i)}(\theta)(\sigma_{abi} + \delta_{ai}\delta_{bi}),
\end{aligned}$$

$I_{(i)}(\theta)$ is the item information function for the 2PL model, and $Z_n(\theta; \mathbf{y}) = 1.7\sum_{i=1}^{n} a_i(y_i - P_i^*(\theta))$.

Let us explain the theorem furthermore. From (17) and (18), we know that $\hat{Z}_n(\hat{\theta}_m; \mathbf{y}) = 0$, and $\hat{Z}_n(\hat{\theta}_w; \mathbf{y}) - \hat{B}_n(\hat{\theta}_w)\hat{I}_n(\hat{\theta}_w) = 0$. Thus, according to (22), the main part of the bias of the naive MLE evaluated at $\hat{\theta}_m$ is $[D_n(\hat{\theta}_m) + J_n(\hat{\theta}_m)]/I_n(\hat{\theta}_m)$. Similarly, according to (23), the main part of the bias of the naive WLE evaluated at $\hat{\theta}_w$ is $[D_n(\hat{\theta}_w) + J_n(\hat{\theta}_w)]/I_n(\hat{\theta}_w)$. Hence,

$$[D_n(\theta) + J_n(\theta)]/I_n(\theta) \tag{27}$$

is the bias function of the naive ability estimators due to the uncertainty of item parameter estimators. It is a function of θ with the biases $\{\delta_{ai}, \delta_{bi}, \delta_{ci}\}$ and covariance matrixes $\{\Sigma_i\}$ of item parameter estimators as its parameters.

The three components of $D_n(\theta)$, $[D_{n,1}(\theta)$, $D_{n,2}(\theta)$, and $D_{n,3}(\theta)]$, divided by $I_n(\theta)$ represent the components of the bias of the naive ability estimators contributed by δ_{ai}, δ_{bi}, and δ_{ci}, respectively. For example,

$$\frac{D_{n,2}(\theta)}{I_n(\theta)} = \sum_{i=1}^{n} \frac{I_{(i)}(\theta)}{I_n(\theta)}\delta_{bi}$$

is the weighted average bias of item difficulty parameters with the proportion of item information as the weight. When most of the biases of item difficulty parameters are in the same direction, $D_{n,2}(\theta)/I_n(\theta)$ tends to be large, and consequently, the bias of the naive ability estimators can also be large. If item parameter estimators are (or assumed to be) unbiased, then all of the δs are zero, and therefore $D_n(\theta)$ is zero in the bias function (27).

The other term, $J_n(\theta)/I_n(\theta)$, in (27) corresponds to the components of the bias caused by the second orders of $\{\delta_{ai},\ \delta_{bi},\ \delta_{ci}\}$, the variances and covariances of item parameter estimators. For example, $J_{n,1}(\theta)/I_n(\theta)$ corresponds to the component of the bias caused by the MSE of $\hat{a}_i$.

It is important to note that when estimated item parameters are treated as true values, $D_n(\theta) \equiv 0$ and $J_n(\theta) \equiv 0$ since all of the δs and σs are zero in this scenario. Under this stronger condition, Warm (1989) proved that BIAS($\hat{\theta}_w$))=$o_p(n^{-1})$ while in the theorem the order of the remainder in (22) or (23) is roughly $o_p(n^{-1/2})$. We get the weaker result under the weaker condition.

The theorem does not have restrictions on the method of item parameter estimation. Hence, any regular joint MLE, marginal MLE, or Bayesian estimation methods can be used to estimate item parameters before applying the theorem. The most widely used IRT model estimation programs at present are BILOG (Mislevy & Bock, 1982) and PARSCALE (Muraki & Bock, 1997). They both use the marginal MLE estimation method.

To calculate the bias function (27), one needs the estimates of $\{\delta_{ai}, \delta_{bi}, \delta_{ci}\}$ and $\{\Sigma_i\}$. Usually, an item parameter estimation program would provide the estimates of the variances and covariances (i.e., $\hat{\Sigma}_i$) along with the item parameter estimates. If an estimation program does not provide estimates of Σ_i, one can always calculate the appropriate information matrix or Hessian matrix to obtain estimates of these covariance matrixes. However, estimates of δ_{ai}, δ_{bi}, and δ_{ci} are typically not directly available from a calibration program. It is a big challenge to estimate, even approximately, the degree of bias because bias cannot be calculated directly from the model. Other techniques, such as the Jackknife method and the Bootstrap method (Efron, 1982), can be used to obtain estimates of δ_{ai}, δ_{bi}, and δ_{ci}. Often, these techniques are developed on an ad hoc basis. A full discussion about bias estimation of item parameter estimators is beyond the scope of this paper.

After $\{\delta_{ai}, \delta_{bi}, \delta_{ci}\}$ and $\{\Sigma_i\}$ are estimated, $D_n(\theta)$ and $J_n(\theta)$ can be estimated by $\hat{D}_n(\theta)$ and $\hat{J}_n(\theta)$, respectively. A caret on a function indicates that all the unknown quantities in the function are replaced by their estimates. Then one can calculate the estimated bias of the naive MLE or the naive WLE of ability using (27) evaluated at $\hat{\theta}_m$ or $\hat{\theta}_w$. One can also determine the range of the bias of the naive ability estimators if the range of the bias, variance, and covariance of item parameter estimators can be obtained. In other words, one can evaluate the impact of measurement errors of item parameter estimators on ability estimation and decide whether the naive MLE or WLE is accurate enough or not. A bias-corrected ability estimator can thus be constructed with the naive WLE $\hat{\theta}_w$ (or MLE $\hat{\theta}_m$) by subtracting the estimated bias function of (27),

$$\hat{\theta}_{wc} = \hat{\theta}_w - [\hat{D}_n(\hat{\theta}_w) + \hat{J}_n(\hat{\theta}_w)]/\hat{I}(\hat{\theta}_w).$$

Below are the steps on how to evaluate the impact of uncertainty about item parameters on the naive ability estimators.

1. Obtain the estimates of variances and covariances of item parameter estimators when estimating item parameters.

2. Estimate the bias of item parameter estimators. The Jackknife, the Bootstrap or any other methods can be used. If one is only interested in the potential impact of uncertainty about item parameters, one may just specify possible amounts (or ranges) of bias of the item parameter estimators. If the bias of item parameter estimators can not be estimated and/or can be negligible, one may assume that the item parameter estimators are unbiased. In this scenario, the simplified formulae (with $D_n(\theta)$ being zero) in the bias function (27) would only reflect the bias of naive ability estimators due to the chance (i.e., random) error of item parameter estimators.

3. Obtain the naive MLE and/or the naive WLE for each test-taker as usual. The ability estimates are usually obtained after the estimation of item parameters in the first step.

4. Calculate the estimated bias function of (27) using the results from the above three steps. Tabulate the bias function at some selected ability levels and draw a plot of the bias against ability level. If the bias is ignorable across the range of θ covering most examinees, the naive MLE or WLE would be acceptable. Otherwise, a bias correction should be applied or some other action is needed, such as using different IRT models to reduce measurement errors of item parameter estimators.

4. A Numerical Example

Suppose that we want to evaluate the bias in a naive ability estimator due to the uncertainty of item parameter estimators that are calibrated from response data of 2,000 examinees and 50 items. For illustration purpose, we use response data from an SAT I: Reasoning Test to demonstrate how to use the procedure proposed in the previous section. The 50 items are the multiple-choice math items. The models used in this study are either 2PL or 3PL models, and PARSCALE is used to estimate item parameters as well as the variances and covariances of item parameter estimators. For each model setting, item parameter estimators are considered either unbiased (i.e., all of the δs are zero) or biased. In the latter case, the bias needs to be estimated.

To estimate the bias of an item parameter estimator, we construct two estimates of the item parameter, one with the (possible) bias and the other without, and then the difference between these two estimates can serve as an estimate of the bias according to model (19), (20), or (21). To obtain unbiased estimates of item parameters, we use the whole data of the SAT administration with 350,400 examinees. Since asymptotically unbiased, theoretically speaking, the estimators from the whole data set can be treated as unbiased for such a large data set. The estimated item parameters from the 2000-examinee sample can serve as the biased ones. However, the difference between biased and unbiased estimates for each item parameter contains not only the bias but also random error. To remove the random error, we need some (e.g., 100) replications of the (possibly biased) estimates of these item parameters. This is done by independently selecting 100 sets of 2,000-examinee simple random samples from the whole data set. These 100 samples are randomly equivalent. We estimate item parameters for each sample, and get 100 sets of (possibly biased) estimated item parameters. Then we calculate the average of the estimates over 100 replications for each item parameter to eliminate or reduce the random error. The difference between the average of the estimates over 100 replications and the corresponding (unbiased) estimate from the whole data set is regarded as an estimate of the bias δ of that item parameter. The above process is done separately for 2PL and 3PL models. Tables 1 and 2 present the bias of item parameter estimators for 2PL and 3PL models, respectively. As shown from Table 1, the bias of item parameter estimators for 2PL models ranges from -0.011 to 0.020 with mean 0.001 for the a parameters and from -0.431 to 0.039 with mean -0.010

for the b parameters. The corresponding quantities for 3PL models shown in Table 2 are noticeably larger: ranging from -2.076 to 0.400 with mean 0.003 for the a parameters, from -0.281 to 4.654 with mean 0.328 for the b parameters, and from -0.314 to 0.319 with mean -0.042 for the c parameters. The sample covariance matrix of $(\hat{a}_i, \hat{b}_i, \hat{c}_i)$ based on the 100 replications can also be calculated and used as a substitute for the estimate of the covariance matrix of $(\hat{a}_i, \hat{b}_i, \hat{c}_i)$. The covariance matrixes are not reported here because their sizes are too large. Note that the variances (or standard errors) of item parameter estimators for 3PL models are also noticeably larger than the corresponding ones for 2PL models.

Insert Tables 1 and 2 about here

Note that the above method for estimating the bias of item parameter estimators is very limited and may not be practical due to the need of a large response data set. As discussed in the previous section, in general, the Bootstrap or the Jackknife method can be used instead to estimate the bias of item parameter estimators.

Insert Table 3 about here

After obtaining the variances, covariances and biases of item parameter estimators, we can calculate the estimated bias function (27) of naive ability estimators at any ability level or at a specific value of an ability estimator (e.g., $\hat{\theta}_w$). In this example, we select 13 ability levels and present the results in Table 3. Columns 2 and 3 of Table 3 show the values of estimated bias function at selected ability levels for 2PL models: The values in column 2 are calculated with the bias estimates of item parameter estimators presented in Table 1, while the values in column 3 are calculated when the item parameter estimators are assumed to be unbiased. In other words, the bias of the naive ability estimators in column 3 is caused only by the variance and covariance (or the chance error) of item parameter estimators, while the bias in column 2 is caused not only by the chance error but also by the bias of item parameter estimators. These results are also illustrated in Figure 1. Columns 4 and 5 of Table 3 present the corresponding results for 3PL models, which are also displayed in Figure 2. Note that the vertical scales in Figures 1 and 2 differ greatly.

Insert Figures 1 and 2 about here

Some interesting results can be observed from this numerical example. First, the bias function in the 2PL case is pretty small at all ability levels between -3.0 to 3.0 while in the 3PL case, it is much larger at most ability levels in the same interval, whether the item parameter estimators are assumed unbiased or not as shown in Figures 1 and 2. Furthermore, the bias in the 3PL case goes in different directions in different θ ranges. These results show that the impact of uncertainty about item parameters on the naive ability estimators (e.g., $\hat{\theta}_w$ satisfying (18)) is very little in the 2PL case to this data set while in the 3PL case, the impact is pretty large. Second, there is no large difference in the bias function between the scenario that all δs are assumed to be zero (i.e., the item parameter estimators are assumed unbiased) and the scenario that the estimates of δs are not zero as presented in Table 1 for the 2PL case or Table 2 for the 3PL case. This result indicates that it is possible that the chance error of item parameter estimators alone can have a strong influence on the naive ability estimators. This observation is important because chance error always exists in item parameter estimators. For this particular data set, we can conclude that the naive MLE or WLE of ability is still appropriate in the 2PL case since the bias function is pretty small (see Figure 1), while in the 3PL case, the naive estimators may not be accurate enough due to the large bias as displayed in Figure 2. Here, however, we do not intend to judge which IRT model (2PL or 3PL) fits the data better.

5. Discussion

In this paper, we derived asymptotic expansion formulae for the naive MLE and WLE of ability using the measurement error approach, and obtained the bias function of the naive ability estimators. Given the estimates of the bias, variance, and covariance of item parameter estimators, the bias function of the naive ability estimators can be calculated. Based on the values of the bias function at ability estimates or targeted ability levels, the impact of uncertainty about item parameters can be evaluated, and consequently, the appropriateness of the naive MLE or WLE can also be appraised. The numerical results presented in the previous section demonstrated one application of the formulae to explore the applicability of the naive ability estimators under different IRT models.

The measurement error models (19), (20) and (21) used in this paper are linear without any distribution requirements for the random errors in the models. Typically, this distribution-free feature guarantees robustness of the approach but loses a degree of efficiency. It will be interesting to see how the approach improves when the distributions of the random errors are specified. A common assumption for the distribution of a random error is a normal distribution. However, the normality assumption is not appropriate for the linear models (19) and (21) because of the limitations of the ranges of the a and c parameters. One way to make the normality assumption reasonable is to use nonlinear measurement error models such as a log linear model for the a parameters and a logit linear model for the c parameters. The measurement error approach can also be applied to any other IRT areas, such as equating and linking, where item parameter estimators are typically treated as fixed without measurement error.

Appendix. Proof of the Theorem

We need two lemmas to prove the theorem. Let

$$S_n(\theta; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \frac{Z_n(\theta)}{n} = \frac{1.7}{n} \sum_{i=1}^{n} a_i K_i(\theta)(y_i - P_i(\theta)).$$

Lemma 1. Suppose that regularity conditions (C0)–(C3) hold. Then,

$$S_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) = \frac{1}{n}[D_n(\theta) + J_n(\theta) + Z_n(\theta)] + o_p\left(\max\left(\sigma_n^2, \frac{1}{\sqrt{n}}\right)\right). \tag{28}$$

Lemma 2. Suppose that regularity conditions (C0)–(C3) hold. For any fixed θ,

$$\hat{I}(\theta) - I(\theta) = o_p(n). \tag{29}$$

Proof of Lemma 1.

$$\begin{aligned} S_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - S_n(\theta; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) &= \frac{1.7}{n} \sum_{i=1}^{n} \hat{a}_i \hat{K}_i(\theta)(y_i - \hat{P}_i(\theta)) - \frac{1.7}{n} \sum_{i=1}^{n} a_i K_i(\theta)(y_i - P_i(\theta)) \\ &= (I) + (II) + (III), \end{aligned} \tag{30}$$

where

$$
\begin{aligned}
(I) &= \frac{1.7}{n}\sum_{i=1}^{n}(\hat{a}_i\hat{K}_i(\theta) - a_iK_i(\theta))(y_i - P_i(\theta)),\\
(II) &= \frac{1.7}{n}\sum_{i=1}^{n} a_iK_i(\theta)(P_i(\theta) - \hat{P}_i(\theta)),\\
(III) &= \frac{1.7}{n}\sum_{i=1}^{n}(\hat{a}_i\hat{K}_i(\theta) - a_iK_i(\theta))(P_i(\theta) - \hat{P}_i(\theta).
\end{aligned}
$$

$$
\begin{aligned}
(I) &= \frac{1.7}{n}\sum_{i=1}^{n}\hat{K}_i(\theta)(\hat{a}_i - a_i)(y_i - P_i(\theta)) + \frac{1.7}{n}\sum_{i=1}^{n} a_i(\hat{K}_i(\theta) - K_i(\theta))(y_i - P_i(\theta))\\
&\triangleq (I_1) + (I_2).
\end{aligned}
$$

By Chebyshev's inequality (see Billingsley, 1995) and condition (C1), we get

$$(I_1) = o_p(1/\sqrt{n}).$$

By the mean value theorem, Chebyshev's inequality and conditions (C0) and (C1), we can obtain

$$(I_2) = o_p(1/\sqrt{n}).$$

Combining above two results we get

$$(I) = o_p(1/\sqrt{n}). \tag{31}$$

By Taylor's theorem and condition (C3),

$$
\begin{aligned}
(II) &= -\frac{1.7}{n}\sum_{i=1}^{n} a_i K_i(\theta)[P_i(\theta;\hat{a}_i,\hat{b}_i,,\hat{c}_i) - P_i(\theta;a_i,b_i,c_i)] \\
&= -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\hat{a}_i-a_i) \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\hat{b}_i-b_i) \\
&\quad -\frac{1.7}{n}\sum_{i=1,c_i>0}^{n} a_i K_i(\theta)Q_i^*(\theta)(\hat{c}_i-c_i) \\
&\quad -\frac{1.7^3}{2n}\sum_{i=1}^{n} a_i(\theta-b_i)^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(1-2P_i^*(\theta)(\hat{a}_i-a_i)^2 \\
&\quad -\frac{1.7^3}{2n}\sum_{i=1}^{n} a_i^3(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(1-2P_i^*(\theta))(\hat{b}_i-b_i)^2 \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta) \\
&\qquad\qquad [1.7a_i(\theta-b_i)(1-2P_i^*(\theta))+1](\hat{a}_i-a_i)(\hat{b}_i-b_i) \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\hat{a}_i-a_i)(\hat{c}_i-c_i) \\
&\quad -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2 K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\hat{b}_i-b_i)(\hat{c}_i-c_i) + o_p(\sigma_n^2).
\end{aligned}
$$

The major terms of (II) can be further decomposed into three parts, (II^*), (II_1) and (II_2). That is,

$$(II) = (II^*) + (II_1) + (II_2) + o_p(\sigma_n^2),$$

where

$$
\begin{aligned}
(II^*) \quad = \quad & -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)\delta_{ai} \\
& +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)\delta_{bi} \\
& -\frac{1.7}{n}\sum_{i=1,c_i>0}^{n} a_iK_i(\theta)Q_i^*(\theta)\delta_{ci} \\
& -\frac{1.7^3}{2n}\sum_{i=1}^{n} a_i(\theta-b_i)^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(1-2P_i^*(\theta))(\sigma_{ai}^2+\delta_{ai}^2) \\
& -\frac{1.7^3}{2n}\sum_{i=1}^{n} a_i^3(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(1-2P_i^*(\theta))(\sigma_{bi}^2+\delta_{bi}^2) \\
& +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)[1.7a_i(\theta-b_i)(1-2P_i^*(\theta))+1](\sigma_{abi}+\delta_{ai}\delta_{bi}) \\
& +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\sigma_{aci}+\delta_{ai}\delta_{ci}) \\
& -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\sigma_{bci}+\delta_{bi}\delta_{ci}) \\
\triangleq \quad & \frac{1}{n}[D_n+J_{n,1}^*+J_{n,2}^*+J_{n,3}^*+J_{n,5}^*+J_{n,6}^*], \qquad (32)
\end{aligned}
$$

(Note that $J_{n,1}^*$, $J_{n,2}^*$, $J_{n,3}^*$, $J_{n,5}^*$ and $J_{n,6}^*$ are parts of $J_{n,1}$, $J_{n,2}$, $J_{n,3}$, $J_{n,5}$ and $J_{n,6}$ defined in the theorem, respectively.)

$$
\begin{aligned}
(II_1) \quad = \quad & -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)[(\hat{a}_i-a_i)-\delta_{ai}] \\
& +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)[(\hat{b}_i-b_i)-\delta_{bi}], \\
& -\frac{1.7}{n}\sum_{i=1,c_i>0}^{n} a_iK_i(\theta)Q_i^*(\theta)[(\hat{c}_i-c_i)-\delta_{ci}],
\end{aligned}
$$

and

$$\begin{aligned}
(II_2) &= -\frac{1.7^3}{2n}\sum_{i=1}^{n} a_i(\theta-b_i)^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(1-2P_i^*(\theta))[(\hat{a}_i-a_i)^2-(\sigma_{ai}^2+\delta_{ai}^2)] \\
&\quad -\frac{1.7^3}{2n}\sum_{i=1}^{n} a_i^3(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)(1-2P_i^*(\theta))[(\hat{b}_i-b_i)^2-(\sigma_{bi}^2+\delta_{bi}^2)] \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)[1,7a_i(\theta-b_i)(1-2P_i^*(\theta))+1] \\
&\qquad\qquad [(\hat{a}_i-a_i)(\hat{b}_i-b_i)-(\sigma_{abi}+\delta_{ai}\delta_{bi})], \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)[(\hat{a}_i-a_i)(\hat{c}_i-c_i)-(\sigma_{aci}+\delta_{ai}\delta_{ci})] \\
&\quad -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2K_i(\theta)P_i^*(\theta)Q_i^*(\theta)[(\hat{b}_i-b_i)(\hat{c}_i-c_i)-(\sigma_{bci}+\delta_{bi}\delta_{ci})].
\end{aligned}$$

By Chebyshev's inequality and conditions (C0) and (C1),

$$\begin{aligned}
(II_1) &= -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i(\theta-b_i)(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)\varepsilon_{ai} \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(1-c_i)K_i(\theta)P_i^*(\theta)Q_i^*(\theta)\varepsilon_{bi}, \\
&\quad -\frac{1.7}{n}\sum_{i=1,c_i>0}^{n} a_iK_i(\theta)Q_i^*(\theta)\varepsilon_{ci} \\
&= o_p(1/\sqrt{n}).
\end{aligned}$$

By Chebyshev's inequality and conditions (C0) and (C2), each of the five terms in (II_2) can be proved to be $o_p(1/\sqrt{n})$; that is,

$$(II_2) = o_p(1/\sqrt{n}).$$

Combining above three parts, we obtain

$$(II) = \frac{1}{n}[D_n + J_{n,1}^* + J_{n,2}^* + J_{n,3}^* + J_{n,5}^* + J_{n,6}^*] + o_p\left(\max\left(\sigma_n^2,\ \frac{1}{\sqrt{n}}\right)\right). \qquad (33)$$

(III) can be decomposed as

$$\begin{aligned}
(III) &= -\frac{1.7}{n}\sum_{i=1}^{n}(\hat{a}_i-a_i)K_i(\theta)[\hat{P}_i(\theta)-P_i(\theta)] - \frac{1.7}{n}\sum_{i=1}^{n} a_i[\hat{K}_i(\theta)-K_i(\theta)][\hat{P}_i(\theta)-P_i(\theta)] \\
&\quad -\frac{1.7}{n}\sum_{i=1}^{n}(\hat{a}_i-a_i)[\hat{K}_i(\theta)-K_i(\theta)][\hat{P}_i(\theta)-P_i(\theta)] \\
&\triangleq (III_1) + (III_2) + o_p(\sigma_n^2).
\end{aligned}$$

By Taylor's theorem,

$$\hat{P}_i(\theta) - P_i(\theta) = G_i(\theta) + o_p(\sigma_n),$$

where

$$G_i(\theta) = 1.7(\theta - b_i)(1-c_i)P_i^*(\theta)Q_i^*(\theta)(\hat{a}_i - a_i) - 1.7a_i(1-c_i)P_i^*(\theta)Q_i^*(\theta)(\hat{b}_i - b_i) + Q_i^*(\theta)(\hat{c}_i - c_i).$$

Thus,

$$(III_1) = -\frac{1.7}{n}\sum_{i=1}^{n}(\hat{a}_i - a_i)K_i(\theta)G_i(\theta) + o_p(\sigma_n^2).$$

Applying the same decomposition technique used above, we can obtain

$$(III_1) = (III_1^*) + o_p\left(\max\left(\sigma_n^2,\ \frac{1}{\sqrt{n}}\right)\right),$$

where

$$\begin{aligned}(III_1^*) &= -\frac{1.7}{n}\sum_{i=1}^{n}K_i(\theta)[1.7(\theta - b_i)(1-c_i)P_i^*(\theta)Q_i^*(\theta)(\sigma_{ai}^2 + \delta_{ai}^2) \\ &\quad - 1.7a_i(1-c_i)P_i^*(\theta)Q_i^*(\theta)(\sigma_{abi} + \delta_{ai}\delta_{bi}) + Q_i^*(\theta)(\sigma_{aci} + \delta_{ai}\delta_{ci})] \\ &\triangleq \frac{1}{n}[J_{n,1}^{**} + J_{n,3}^{**} + J_{n,5}^{**}]. \qquad (34)\end{aligned}$$

Here $J_{n,1}^{**}$, $J_{n,3}^{**}$ and $J_{n,5}^{**}$ are parts of $J_{n,1}$, $J_{n,3}$ and $J_{n,5}$, respectively.

Similarly, we can obtain

$$(III_2) = (III_2^*) + o_p\left(\max\left(\sigma^2,\ \frac{1}{\sqrt{n}}\right)\right),$$

where

$$\begin{aligned}(III_2^*) &= -\frac{1.7^3}{n}\sum_{i=1}^{n}a_i(\theta - b_i)^2c_i(1-c_i)K_i(\theta)L_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\sigma_{ai}^2 + \delta_{ai}^2) \\ &\quad -\frac{1.7^3}{n}\sum_{i=1}^{n}a_i^3c_i(1-c_i)K_i(\theta)L_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\sigma_{bi}^2 + \delta_{bi}^2) \\ &\quad +\frac{1.7^3}{n}\sum_{i=1}^{n}2a_i^2(\theta - b_i)c_i(1-c_i)K_i(\theta)L_i(\theta)P_i^*(\theta)Q_i^*(\theta)(\sigma_{abi} + \delta_{ai}\delta_{bi}) \\ &\quad +\frac{1.7}{n}\sum_{i=1}^{n}a_iK_i(\theta)L_i(\theta)Q_i^*(\theta)(\sigma_{ci}^2 + \delta_{ci}^2) \\ &\quad -\frac{1.7^2}{n}\sum_{i=1}^{n}a_i(\theta - b_i)K_i(\theta)L_i(\theta)Q_i^*(\theta)[c_i - (1-c_i)P_i^*(\theta)](\sigma_{aci} + \delta_{ai}\delta_{ci}) \\ &\quad +\frac{1.7^2}{n}\sum_{i=1}^{n}a_i^2K_i(\theta)L_i(\theta)Q_i^*(\theta)[c_i - (1-c_i)P_i^*(\theta)](\sigma_{bci} + \delta_{bi}\delta_{ci}) \\ &\triangleq \frac{1}{n}[J_{n,1}^{***} + J_{n,2}^{***} + J_{n,3}^{***} + J_{n,4}^{***} + J_{n,5}^{***} + J_{n,6}^{***}]. \qquad (35)\end{aligned}$$

Here $J^{***}_{n,1}, \ldots, J^{***}_{n,6}$ are parts of $J_{n,1}, \ldots, J_{n,6}$, respectively. Thus,

$$
\begin{aligned}
(III) &= \frac{1}{n}[(J^{**}_{n,1} + J^{***}_{n,1}) + J^{***}_{n,2} + (J^{**}_{n,3} + J^{***}_{n,3}) + J^{***}_{n,4} + (J^{**}_{n,5} + J^{***}_{n,5}) + J^{***}_{n,6}] \\
&\quad + o_p\left(\max\left(\sigma^2, \ \frac{1}{\sqrt{n}}\right)\right). \qquad (36)
\end{aligned}
$$

From (32), (34) and (35) and according to the notation in the theorem, it can be verified that $J^{*}_{n,1} + J^{**}_{n,1} + J^{***}_{n,1} = J_{n,1}$, $J^{*}_{n,2} + J^{***}_{n,2} = J_{n,3}$, $J^{*}_{n,3} + J^{**}_{n,3} + J^{***}_{n,3} = J_{n,3}$, $J^{***}_{n,4} = J_{n,4}$, $J^{*}_{n,5} + J^{**}_{n,5} + J^{***}_{n,5} = J_{n,5}$, and $J^{*}_{n,6} + J^{***}_{n,6} = J_{n,6}$. By (30), (31), (33) and (36), we can obtain (28). □

Proof of Lemma 2. Using the same technique as in the proof of Lemma 1,

$$
\begin{aligned}
&\frac{1}{n}[I_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - I_n(\theta; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})] \\
&= \frac{1.7^2}{n}\sum_{i=1}^{n}[\hat{a}_i^2(1-\hat{c}_i)\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - a_i^2(1-c_i)P^*_i(\theta)Q^*_i(\theta)K_i(\theta)] \\
&= \frac{1.7^2}{n}\sum_{i=1}^{n}\{(\hat{a}_i^2 - a_i^2)(1-\hat{c}_i)\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - a_i^2(\hat{c}_i - c_i)\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) \\
&\qquad\qquad + a_i^2(1-c_i)[\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - P^*_i(\theta)Q^*_i(\theta)K_i(\theta)]\} \\
&= \frac{1.7^2}{n}\sum_{i=1}^{n}[(\hat{a}_i - a_i)^2 + 2a_i(\hat{a}_i - a_i)](1-\hat{c}_i)\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) \\
&\quad -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(\hat{c}_i - c_i)P^*_i(\theta)Q^*_i(\theta)K_i(\theta) \\
&\quad -\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(\hat{c}_i - c_i)[\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - P^*_i(\theta)Q^*_i(\theta)K_i(\theta)] \\
&\quad +\frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(1-c_i)[\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - P^*_i(\theta)Q^*_i(\theta)K_i(\theta)] \\
&= \frac{1.7^2}{n}\sum_{i=1}^{n} a_i^2(1-c_i)[\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - P^*_i(\theta)Q^*_i(\theta)K_i(\theta)] + o_p(1).
\end{aligned}
$$

By Taylor's theorem,

$$
\begin{aligned}
&\hat{P}^*_i(\theta)\hat{Q}^*_i(\theta)\hat{K}_i(\theta) - P^*_i(\theta)Q^*_i(\theta)K_i(\theta) = \\
&\qquad P^*_i(\theta)Q^*_i(\theta)K_i(\theta)\{1.7(\theta - b_i)[1 - 2P^*_i(\theta) + c_iL_i(\theta)](\hat{a}_i - a_i) \\
&\qquad\qquad - 1.7a_i[1 - 2P^*_i(\theta) + c_iL_i(\theta)](\hat{b}_i - b_i) - L_i(\theta)(\hat{c}_i - c_i)\} + o_p(\sigma_n).
\end{aligned}
$$

By the regularity conditions, we obtain (29). ∎

Proof of Theorem. From (18),

$$S_n(\hat{\theta}_w; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - \frac{\hat{B}(\hat{\theta}_w)\hat{I}(\hat{\theta}_w)}{n} = 0.$$

Thus,

$$\begin{aligned} S_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - \frac{B(\theta)I(\theta)}{n} &= S_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - S_n(\hat{\theta}_w; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) + \frac{1}{n}[\hat{B}(\hat{\theta}_w)\hat{I}(\hat{\theta}_w) - B(\theta)I(\theta)] \\ &= \frac{1.7}{n}\sum_{i=1}^{n} \hat{a}_i\{\hat{K}_i(\hat{\theta}_w)[\hat{P}_i(\hat{\theta}_w) - y_i] - \hat{K}_i(\theta)[\hat{P}_i(\theta) - y_i]\} + U_n, \end{aligned}$$

where

$$U_n = \frac{1}{n}[\hat{B}(\hat{\theta}_w)\hat{I}(\hat{\theta}_w) - B(\theta)I(\theta)].$$

By the mean value theorem, there exists a point $\bar{\theta}$ between θ and $\hat{\theta}_w$ such that

$$\begin{aligned} & S_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - \frac{B(\theta)I(\theta)}{n} \\ &= \frac{1.7^2}{n}\sum_{i=1}^{n} \hat{a}_i^2\hat{K}_i(\bar{\theta})\{(1-\hat{c}_i)\hat{P}_i^*(\bar{\theta})\hat{Q}_i^*(\bar{\theta}) + \hat{c}_i\hat{L}_i(\bar{\theta})[\hat{P}_i(\bar{\theta}) - y_i]\}(\hat{\theta} - \theta) + U_n \\ &= \frac{1}{n}\hat{I}(\bar{\theta})(\hat{\theta}_w - \theta) + T_n(\hat{\theta}_w - \theta) + U_n, \end{aligned} \tag{37}$$

where

$$T_n = \frac{1.7^2}{n}\sum_{i=1}^{n} \hat{a}_i^2\hat{c}_i\hat{K}_i(\bar{\theta})\hat{L}_i(\bar{\theta})[\hat{P}_i(\bar{\theta}) - y_i].$$

It can be verified that $T_n = o_p(1/\sqrt{n})$ and $U_n = o_p(1/\sqrt{n})$. Thus, from (37) and by Lemma 2, we obtain

$$\begin{aligned} S_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - \frac{B(\theta)I(\theta)}{n} &= \frac{1}{n}I(\bar{\theta})(\hat{\theta}_w - \theta) + o_p(1/\sqrt{n}) \\ &= \frac{1}{n}I(\theta)(\hat{\theta}_w - \theta) + \frac{1}{n}[I(\bar{\theta}) - I(\theta)](\hat{\theta}_w - \theta) + o_p(1/\sqrt{n}) \\ &= \frac{1}{n}I(\theta)(\hat{\theta}_w - \theta) + o_p(1/\sqrt{n}). \end{aligned}$$

By condition (C4),

$$\hat{\theta} - \theta = \frac{nS_n(\theta; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) - B(\theta)I(\theta)}{I(\theta)} + o_p(1/\sqrt{n}).$$

By Lemma 1, we have (23).

Similarly, we can prove (22). ∎

References

Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York: Wiley.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd). London: Chapman & Hall.

Chang, H., & Stout, W. F. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika, 58*, 37–52.

Efron, B. (1982). *The Jackknife, the Bootstrap, and other resampling plans.* Philadelphia, PA: Society for Industrial and Applied mathematics.

Fuller, W. A. (1987). *Measurement Error Models.* New York: John Wiley.

Lewis, C. (1985, June). *Estimating individual abilities with imperfectly known item response functions.* Paper presented at the annual meeting of the Psychometric Society, Nashville, TN.

Lewis, C. (2001). Expected response functions. In A. Boomsma, M. van Duijin, & T. Snijders (Eds.), *Essays on item response theory* (pp. 163–171). New York: Springer-Verlag.

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.

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

Mislevy, R. J., Wingersky, M. S., & Sheehan, K. M. (1994). *Dealing with uncertainty about item parameters: Expected response functions* (ETS Research Report 94-28-ONR). Princeton, NJ: Educational Testing Service.

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

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* New York, NY: John Wiley & Sons.

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, New Brunswick, NJ.

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

Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika, 55*, 371–390.

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

Table 1. Bias Estimates ($\hat{\delta}$) of Item Parameter Estimators of 3PL Models With Sample Size 2,000

Item	a	b	Item	a	b
1	-0.006	-0.018	26	0.002	-0.007
2	0.020	-0.101	27	0.013	0.008
3	0.018	-0.005	28	-0.002	0.001
4	0.012	-0.007	29	0.004	-0.004
5	0.001	-0.003	30	-0.003	-0.008
6	0.010	0.001	31	0.004	-0.001
7	-0.004	-0.007	32	0.003	0.003
8	-0.001	-0.003	33	-0.002	0.003
9	0.004	0.014	34	-0.002	0.001
10	-0.003	-0.007	35	0.004	-0.002
11	0.004	-0.015	36	-0.002	0.007
12	-0.003	-0.002	37	0.004	-0.009
13	0.005	-0.006	38	-0.000	0.004
14	0.009	0.006	39	-0.001	0.006
15	0.001	0.004	40	-0.003	0.018
16	0.005	0.004	41	-0.011	-0.431
17	-0.002	0.003	42	-0.008	-0.041
18	0.002	0.003	43	0.001	-0.007
19	0.001	-0.003	44	0.000	0.016
20	-0.004	0.002	45	0.001	0.001
21	0.004	-0.004	46	0.000	-0.001
22	-0.007	0.020	47	0.001	0.007
23	-0.006	0.007	48	-0.001	-0.002
24	-0.007	0.010	49	0.004	0.014
25	-0.005	0.039	50	0.007	0.005

Table 2. Bias Estimates ($\hat{\delta}$) of Item Parameter Estimators of 3PL Models With Sample Size 2,000

Item	a	b	c	Item	a	b	c
1	-0.806	4.654	0.085	26	0.072	0.119	-0.087
2	-2.076	1.764	0.269	27	0.070	0.064	-0.175
3	0.400	0.275	-0.282	28	0.299	-0.281	-0.314
4	0.099	0.215	-0.177	29	0.050	0.050	-0.139
5	0.106	0.170	-0.026	30	0.115	0.270	-0.097
6	0.115	0.265	-0.106	31	0.103	0.205	-0.025
7	0.058	0.130	-0.099	32	0.105	0.249	-0.049
8	0.141	0.203	-0.007	33	0.077	0.192	-0.016
9	0.069	0.187	-0.033	34	0.073	0.183	-0.025
10	0.028	0.028	-0.080	35	0.052	0.146	-0.042
11	0.060	0.110	-0.090	36	0.060	0.156	-0.016
12	0.092	0.246	-0.091	37	0.037	0.085	-0.052
13	0.127	0.247	-0.019	38	0.091	0.187	-0.011
14	0.103	0.213	-0.022	39	0.091	0.140	-0.002
15	0.061	0.163	-0.020	40	0.081	0.121	-0.001
16	0.079	0.178	-0.015	41	-1.371	3.011	0.319
17	0.086	0.197	-0.014	42	0.137	-0.000	-0.155
18	0.077	0.186	-0.018	43	0.061	-0.039	-0.143
19	0.080	0.204	-0.052	44	0.033	0.043	-0.082
20	0.043	0.150	-0.016	45	0.066	0.175	-0.067
21	0.094	0.127	-0.003	46	0.085	0.207	-0.028
22	0.114	0.152	-0.002	47	0.033	0.109	-0.027
23	0.106	0.172	-0.003	48	0.057	0.173	-0.029
24	0.075	0.158	-0.003	49	0.073	0.166	-0.009
25	0.164	0.126	-0.000	50	0.103	0.124	-0.003

Table 3. Bias of Naive Ability Estimators Due to Uncertainty About Item Parameters

	2PL		3PL	
Ability	$\hat{\delta}$	$\hat{\delta} \equiv 0$	$\hat{\delta}$	$\hat{\delta} \equiv 0$
-3.0	0.031	0.020	5.736	4.974
-2.5	0.018	0.011	3.043	2.480
-2.0	0.012	0.007	1.918	1.142
-1.5	0.008	0.005	1.273	0.490
-1.0	0.006	0.003	0.837	0.193
-0.5	0.003	0.002	0.509	0.043
0.0	0.001	0.001	0.261	-0.044
0.5	-0.001	0.000	0.078	-0.108
1.0	-0.002	0.000	-0.083	-0.174
1.5	-0.003	0.001	-0.278	-0.273
2.0	-0.003	0.003	-0.704	-0.519
2.5	-0.001	0.006	-1.909	-1.256
3.0	0.001	0.010	-5.375	-3.440

Figure 1. Bias Functions of Naive Ability Estimators in 2PL Model: Item Parameter Estimators Assumed Biased or Not

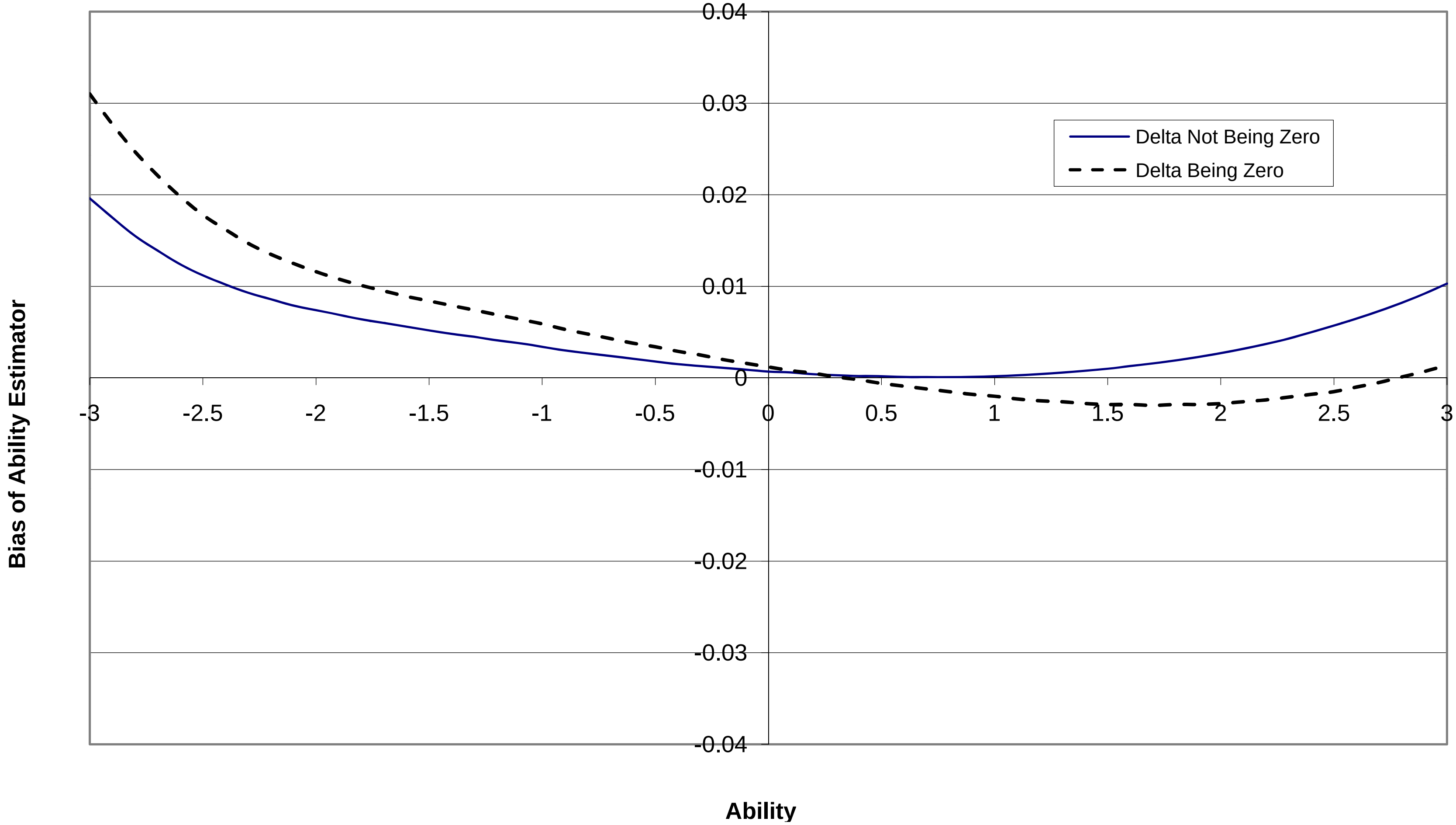

Figure 2. Bias Functions of Naive Ability Estimators in 3PL Model: Item Parameter Estimators Assumed Biased or Not