Research Report

Asymptotic Limits of Item Parameters in Joint Maximum-Likelihood Estimation for the Rasch Model

Shelby J. Haberman

February 2008

ETS RR-08-04

Asymptotic Limits of Item Parameters in Joint Maximum-Likelihood Estimation for the Rasch Model

Shelby J. Haberman
ETS, Princeton, NJ

February 2008

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

Techniques are developed for approximation and exact computation of the asymptotic limit of the item parameter estimates obtained by application of joint maximum-likelihood estimation to the Rasch model.

Key words: Asymptotic bias, logit analysis

Acknowledgements

This paper has benefited from many discussions with Matthias von Davier. The author thanks Yi-Hsuan Lee and Dan Eignor for helpful comments.

For the Rasch model (Rasch, 1960) for binary responses, estimation of person and item parameters via joint maximum likelihood remains common despite consistency problems that have been known for a long period of time (Andersen, 1973) and despite the fact that these consistency problems disappear if conditional maximum likelihood is employed. To some extent, this practice appears to reflect the existence of American commercial software such as Winsteps that performs joint estimation, although Winmira and Conquest are example of readily available commercial software for conditional maximum likelihood. To some extent, the user of joint estimation may be influenced by the fact that bias problems decrease as the number of items becomes large (Haberman, 1977, 2004). In this report, some tools are provided for bias assessment.

Section 1 summarizes known results concerning bias. Section 2 provides an approximation for bias in the case in which the variability of item parameters is small. Section 3 provides a general approach to computation of bias for a given set of item parameters and a given ability distribution. Section 4 considers consequences of the results of this report when joint estimation is applied to equating. Section 5 provides some concluding observations.

1 Asymptotic Limits for Item Parameters

In this section, the basic limiting behavior of maximum-likelihood estimates is considered for the binary Rasch model (Andersen, 1973; Fischer, 1981; Haberman, 1977, 2004). Results in this section are all known. Let X_{ij}, $1 \le i \le n$, $1 \le j \le q$, be binary random variables with values 0 or 1, such that X_{ij} represents a response of an examinee i to an item j, with X_{ij} equal to 1 for a correct response and equal to 0 otherwise. Let $q \ge 2$, let $\mathbf{X}_i$ be the q-dimensional vector of X_{ij}, $1 \le j \le q$, and let θ_i, $1 \le i \le n$, be associated real random variables that are independent and identically distributed. Let the pairs $(\mathbf{X}_i, \theta_i)$, $1 \le i \le n$, be mutually independent, and let the X_{ij}, $1 \le j \le q$, be conditionally independent given θ_i. For real x and y, let $P(x,y) = [1+\exp(y-x)]^{-1}$. Let p^S_k, $0 \le k \le q$, be the probability that the examinee sum $S_i = \sum_{j=1}^q X_{ij}$ is k, let p_j, $1 \le j \le q$, be the probability that $X_{ij} = 1$, and let m_{kj}, $0 \le k \le q$, $1 \le j \le q$, be the conditional probability that response $X_{ij} = 1$ given that the sum $S_i = k$. The distinct feature of the Rasch model is that, for unknown real parameters β_j, $1 \le j \le q$, the conditional probability that $X_{ij} = 1$ given θ_i is $P(\theta_i, \beta_j)$. To permit parameter identification, let β_1 be assumed to be 0. To assist in joint maximum-likelihood estimation, it is helpful to consider infinite numbers. Adopt the convention that $1/[1+\exp(z)]$ is 1 for $z = -\infty$ and 0 for $z = \infty$. Let $y - x$ be ∞ for $y = \infty$ and $x < \infty$ or for

$y > -\infty$ and $x = -\infty$, let $y - x$ be $-\infty$ for $y = -\infty$ and $x > -\infty$ or for $y < \infty$ and $x = -\infty$, and let $y - x = 0$ if $x = y = \infty$ or $x = y = -\infty$.

In joint maximum-likelihood estimation, estimation proceeds as if the θ_i were fixed parameters. The joint log likelihood function

$$\ell(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} \sum_{j=1}^{q} \{X_{ij} \log P(a_i, b_j) + (1 - X_{ij}) \log[1 - P(a_i, b_j)]\}$$

for $\mathbf{a}$, an n-dimensional extended real vector with coordinates a_i, $1 \le i \le n$, and $\mathbf{b}$, a q-dimensional extended real vector with coordinates b_j, $1 \le j \le q$, $b_1 = 0$. The convention is used that $0 \log 0 = 0$. Let ℓ_M be the supremum of $\ell(\mathbf{a}, \mathbf{b})$ for n-dimensional extended real $\mathbf{a}$ and q-dimensional extended real $\mathbf{b}$ such that $b_1 = 0$, and let J_M be the set of pairs $(\mathbf{a}, \mathbf{b})$ such that $\ell(\mathbf{a}, \mathbf{b}) = \ell_M$.

One may then define extended maximum-likelihood estimates $\hat{\boldsymbol{\theta}}$ of the ability vector $\boldsymbol{\theta}$ with coordinates the examinee abilities θ_i, $1 \le i \le n$, and extended maximum-likelihood estimates $\hat{\boldsymbol{\beta}}$ of the vector $\boldsymbol{\beta}$ with coordinates the item difficulties β_j, $1 \le j \le q$, so that $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$ is in J_M if J_M is nonempty, the initial coordinate $\hat{\beta}_1$ is 0, and $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ are determined by the observed n by q data matrix $\mathbf{X}$ with row i and column j equal to X_{ij}. Note that if J_M is nonempty, $b_1 = 0$, and $\ell(\mathbf{a}, \mathbf{b}) = \ell_M$, then $\hat{\boldsymbol{\theta}} = \mathbf{a}$ and $\hat{\boldsymbol{\beta}} = \mathbf{b}$.

As the sample size n increases, $\hat{\boldsymbol{\beta}}$ converges with probability 1 to a real vector $\boldsymbol{\gamma}$ with coordinates γ_j, $1 \le j \le q$, such that $\gamma_1 = 0$. For some unique extended real θ_{kS}, $0 \le k \le q$,

$$\sum_{k=0}^{q} p_k^S P(\theta_{kS}, \gamma_j) = p_j,\ 1 \le j \le q, \tag{1}$$

and

$$\sum_{j=1}^{q} P(\theta_{kS}, \gamma_j) = k,\ 0 \le k \le q. \tag{2}$$

Equations (1) and (2), together with the constraint that $\gamma_1 = 0$, uniquely determine $\boldsymbol{\gamma}$. By (2), $\theta_{0S} = -\infty$, $\theta_{qS} = \infty$, and θ_{kS} is finite for $1 \le k \le q - 1$.

The basic challenge with joint estimation is that $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ need not be the same, so that joint estimation is asymptotically biased. For example, if $q = 2$, then $\boldsymbol{\gamma} = 2\boldsymbol{\beta}$ (Andersen, 1973, pp. 66–69). Whenever $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ differ, an inconsistency problem exists. The size of the inconsistency is considered in sections 2 and 3.

It should be emphasized that in conditional maximum-likelihood estimation, the problem of asymptotic bias is not present. In conditional likelihood, inferences are conditional on the total score S_i. If Γ is the set of q-dimensional vectors $\mathbf{x}$ with coordinates x_j equal to 0 or 1 for $1 \le j \le q$ and if $\Gamma(k)$, $0 \le k \le q$, is the set of $\mathbf{x}$ in Γ such that the sum of the coordinates is $\sum_{j=1}^q x_j = k$, then the conditional likelihood function $\ell_C(\mathbf{b})$ is defined for q-dimensional real vectors $\mathbf{b}$ such that $b_1 = 0$ by

$$\ell_C(\mathbf{b}) = \sum_{i=1}^n \left[-\sum_{j=1}^q b_j X_{ij} - \log T(\mathbf{b}, S_i) \right],$$

where

$$T(\mathbf{b}, k) = \sum_{\mathbf{x} \in \Gamma(k)} \exp \left[-\sum_{j=1}^q b_j x_j \right]$$

for $0 \le k \le q$. Let the supremum of ℓ_C be ℓ_{CM}, and let J_C be the set of q-dimensional real vectors $\mathbf{b}$ such that $b_1 = 0$ and $\ell_C(\mathbf{b}) = \ell_{CM}$. Let the conditional maximum-likelihood estimate $\hat{\boldsymbol{\beta}}_C$ be a function of the observed X_{ij}, $1 \le i \le n$, $1 \le j \le q$, such that $\ell_C(\hat{\boldsymbol{\beta}}_C) = \ell_{CM}$ whenever J_C is nonempty. Let coordinate j of $\hat{\boldsymbol{\beta}}_C$ be $\hat{\beta}_{jC}$, $1 \le j \le q$. Then $\hat{\boldsymbol{\beta}}_C$ converges with probability 1 to $\boldsymbol{\beta}$ (Andersen, 1973, chapter 5). Extended real versions of $\hat{\boldsymbol{\beta}}_C$ can be considered, but they are relatively complicated to describe and of less practical importance than in joint maximum likelihood. Consequently they are not considered here.

2 The Case of Nearly Equal Item Parameters

Some basic analysis may be conducted by consideration of the case of nearly equal item parameters. One fixes the distribution of the θ_i and lets the vector $\boldsymbol{\beta}$ approach the vector $\mathbf{0}$ with all q coordinates equal to 0. The implicit function theorem is then applied to (1) and (2). The following result is obtained for the maximum norm $|\mathbf{b}| = \max_{1 \le j \le q} |b_j|$ defined for n-dimensional real vectors $\mathbf{b}$.

Theorem 1 *For each real $\delta > 0$ a real $\epsilon > 0$ exists such that $|\gamma_j - q\beta_j/(q-1)| \le \delta|\boldsymbol{\beta}|$ whenever $|\boldsymbol{\beta}| < \epsilon$.*

This theorem, which is proven in the appendix, provides some formal basis for attempts to correct bias in $\hat{\beta}_j$ by multiplication by $(q-1)/q$ (Jansen, Wollenberg, & Wierda, 1988; Wollenberg, Wierda, & Jansen, 1988; Wright, 1988; Wright & Douglas, 1977). For $q = 2$, the result holds with $\delta = 0$ and ϵ, an arbitrary real number. The result is consistent with the general observation that, for any integer $q \ge 2$, if θ_i is a bounded random variable, then, for any real constant $c > 0$, a real $d > 0$ exists such that $|\boldsymbol{\gamma} - \boldsymbol{\beta}| < d/q$ whenever $|\boldsymbol{\beta}| < c$ (Haberman, 2004).

3 Computation of Asymptotic Limits

For a given distribution of θ_i and for a given $\boldsymbol{\beta}$, the asymptotic limit $\boldsymbol{\gamma}$ can be computed with little difficulty by exploiting some techniques for efficient computation of probabilities of sums of independent Bernoulli random variables (Haberman, 2004). Computations also exploit standard methods to calculate maximum-likelihood estimates for logit models. The basic observations are the following. The value of $\boldsymbol{\gamma}$ can be found by maximizing the expected log likelihood $E(\ell(\mathbf{a}, \mathbf{b}))$ for extended real vectors $\mathbf{a}$ of dimension n and extended real vectors $\mathbf{b}$ of dimension q subject to the constraint that $b_1 = 0$ and, for some a'_k, $0 \le k \le q$, $a_i = a'_k$ for any examinee i for whom $S_i = k$. Given that $a'_0 = -\infty$ and $a'_q = \infty$,

$$E(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{q-1} p_k^S [m_{kj} \log P(a'_k, b_j) + (1 - m_{kj}) \log[1 - P(a'_k, b_j)].$$

The maximum is achieved for $a'_k = \theta_{Sk}$ and $b_j = \gamma_j$, subject to the constraint that $\gamma_1 = 0$. This maximization can be accomplished by any standard computer package for calculation of maximum-likelihood estimates for logit models. For each real a, define independent Bernoulli random variables $U_j(a)$, $1 \le j \le q$, so that $U_j(a) = 1$ with probability $P(a, \beta_j)$. Then the product $p_k^S m_{kj}$ is the expectation of $g_{kj}(\theta_i)$, where $g_{kj}(a)$, a real, is $P(a, \beta_j)$ times the probability that $\sum_{h=1, h\ne j}^{q} U_h(a) = k - 1$. Similarly, p_k^S is the expectation of $h_k(\theta_i)$, where $h_k(a)$, a real, is the probability that $\sum_{j=1}^{q} U_j(a) = k$. It follows that $p_k^S(1 - m_{kj})$ is the expectation of $h_k(\theta_i) - g_{kj}(\theta_i)$. Computation of $g_{kj}(a)$ and $h_k(a)$ can be achieved by a recursive algorithm (Haberman, 2004).

A Fortran 95 program was constructed to find values of $\boldsymbol{\gamma}$. Several examples are helpful to illustrate application of Theorem 1. If $q = 5$ and θ_i is 1, 2, or 3 with respective probabilities 0.3, 0.4, and 0.5 and if $\beta_j = j - 1$ for $1 \le j \le 5$, then Theorem 1 suggests that $\gamma_j = 1.25(j - 1)$, so that $\gamma_1 = 0$, $\gamma_2 = 1.25$, $\gamma_3 = 2.5$, $\gamma_4 = 3.75$, and $\gamma_5 = 5$. In fact $\gamma_1 = 0$, $\gamma_2 = 1.451$, $\gamma_3 = 2.793$, $\gamma_4 = 4.075$, and $\gamma_5 = 5.295$. Thus the approximation of Theorem 1 is only moderately accurate in this case, which involves β_j that are not close to 0. The approximation is relatively accurate if $\beta_j = (j - 1)/10$, for the predicted values are $\gamma_1 = 0$, $\gamma_2 = 0.125$, $\gamma_3 = 0.25$, $\gamma_4 = 0.375$, and $\gamma_5 = 0.5$. Actual values are $\gamma_1 = 0$, $\gamma_2 = 0.124$, $\gamma_3 = 0.249$, $\gamma_4 = 0.374$, and $\gamma_5 = 0.500$. Thus the theorem can be helpful with small $|\boldsymbol{\beta}|$, and the theorem is not especially helpful with large $|\boldsymbol{\beta}|$.

The accuracy of the approximation suggested by the theorem can be somewhat worse than in the previous examples if the range of the β_j is increased. Consider the same distribution of θ_i for $\beta_2 = -4$, $\beta_3 = -2$, $\beta_4 = 2$, and $\beta_5 = 4$. In this case, $\gamma_2 = -5$, $\gamma_3 = -2.5$, $\gamma_4 = 2.5$, and $\gamma_5 = 5$ is anticipated, but the actual values are $\gamma_2 = -5.472$, $\gamma_3 = -3.013$, $\gamma_4 = 3.263$, and $\gamma_5 = 6.382$. Although the theorem may not provide a fully satisfactory approximation, it should be emphasized that $\boldsymbol{\gamma}$ is certainly poorly approximated by $\boldsymbol{\beta}$ in numerous cases. For instance, in the last example, $|\boldsymbol{\gamma} - 1.25\boldsymbol{\beta}|$ is 1.382, but $|\boldsymbol{\gamma} - \boldsymbol{\beta}|$ is 2.382.

Although it is clearly helpful for the number of items to increase, it should be emphasized that the difference between $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ can be large for numbers of items encountered in tests. For the approximation of Theorem 1, accuracy results for the same distribution of θ_i but for $q = 16$ and $\beta_j = 0.3(j-1)$ are relatively satisfactory but certainly not precise, for $|\boldsymbol{\gamma} - (16/15)\boldsymbol{\beta}|$ is 0.082 and $|\boldsymbol{\gamma} - \boldsymbol{\beta}|$ is 0.346. If q is 41, if θ_i is -2, 0, or 2 with respective probabilities 0.3, 0.4, and 0.3, if $\beta_j = 0.2(j-22)$ for $2 \le j \le 21$, and if $\beta_j = 0.2(j-21)$ for $22 \le j \le 41$, so that β_j are placed on evenly spaced points from -4 to 4, then $|\boldsymbol{\gamma} - 1.025\boldsymbol{\beta}|$ is 0.054 and $|\boldsymbol{\gamma} - \boldsymbol{\beta}|$ is 0.155. Numerous alternative cases can be considered by variation of the distribution of θ_i, variation of the number q of items, or variation of the item difficulties β_j. Particularly severe problems can be found for θ_i equal to 5 with probability 1, β_j equal 0 for $j \le q/2$ and $\beta_j = 10$ otherwise. For instance, for $q = 20$, $|\boldsymbol{\gamma} - \boldsymbol{\beta}|$ is 4.883 and $|\boldsymbol{\gamma} - (20/19)\boldsymbol{\beta}|$ is 4.356.

4 Joint Maximum Likelihood and Equating

To interpret results in sections 2 and 3, the practical effect of bias may be considered in terms of equating. For this purpose, consider equating of the results for the X_{ij} to a reference form with $q' \ge 2$ items and n' examinees. Let items 1 to $r \ge 1$ be common, where r is less than both q and q'. Let X'_{ij}, $1 \le i \le n'$, $1 \le j \le q'$, be binary random variables with values 0 or 1, such that X'_{ij} represents a response on the reference form of an Examinee i to an Item j, with X'_{ij} equal to 1 for a correct response and equal to 0 otherwise. Assume that the examinees for the two forms are distinct. Let $\mathbf{X}'_i$ be the q'-dimensional vector of X'_{ij}, $1 \le j \le q'$, and let θ'_i, $1 \le i \le n'$, be associated real random variables that are independent and identically distributed. Let the pairs $(\mathbf{X}'_i, \theta'_i)$, $1 \le i \le n$, be mutually independent, and let the X'_{ij}, $1 \le j \le q'$, be conditionally independent given θ_i. For unknown real parameters β'_j, $1 \le j \le q'$, let the conditional probability that $X'_{ij} = 1$ given θ'_i be $P(\theta'_i, \beta'_j)$, and let $\beta'_1 = 0$. Note that if the common items perform as common items, then it should be true that $\beta_j = \beta'_j$ for $1 \le j \le r$, but the analysis here only will use a restraint that the arithmetic mean μ of the β_j, $1 \le j \le r$, is the same as the arithmetic mean μ' of the β'_j, $1 \le j \le r$. Other approaches can be considered that lead to relatively similar results. The approach selected is the simplest to apply without changing procedures for parameter estimation in some fashion.

Let the joint maximum-likelihood estimate of the vector $\boldsymbol{\beta}'$ of β_j', $1 \le j \le q'$, be $\hat{\boldsymbol{\beta}}'$, and let the conditional maximum-likelihood estimate of $\boldsymbol{\beta}'$ be $\hat{\boldsymbol{\beta}}_C'$, so that, as n' becomes large, $\hat{\boldsymbol{\beta}}'$ converges with probability 1 to a limit $\boldsymbol{\gamma}'$ and $\hat{\boldsymbol{\beta}}_C'$ converges with probability 1 to $\boldsymbol{\beta}'$. If $\hat{\mu}$ denotes the arithmetic mean of the $\hat{\beta}_j$ for $1 \le j \le r$, $\hat{\mu}'$ denotes the arithmetic mean of the $\hat{\beta}_j'$ for $1 \le j \le r$, $\hat{\mu}_C$ denotes the arithmetic mean of the $\hat{\beta}_{jC}$ for $1 \le j \le r$, $\hat{\mu}_C'$ denotes the arithmetic mean of the $\hat{\beta}_{jC}'$ for $1 \le j \le r$, ν denotes the arithmetic mean of the γ_j, $1 \le j \le r$, and ν' denotes the arithmetic mean of the γ_j', $1 \le j \le r$, then the probability is 1 that $\hat{\mu}$ converges to ν, $\hat{\mu}'$ converges to ν', $\hat{\mu}_C$ converges to μ, and $\hat{\mu}_C'$ converges to $\mu' = \mu$.

Consider equating by true scores. For the vector $\mathbf{b}$ of item parameters associated with the X_{ij}, the test characteristic curve is

$$V(a, \mathbf{b}) = \sum_{j=1}^{q} P(a, b_j).$$

Similarly, for the vector $\mathbf{b}'$ of item parameters associated with the X_{ij}', the test characteristic curve is

$$V'(a, \mathbf{b}') = \sum_{j=1}^{q'} P(a, b_j').$$

For $\mathbf{b}$ and $\mathbf{b}'$ real vectors, both $q^{-1}V(a, \mathbf{b})$ and $(q')^{-1}V'(a, \mathbf{b})$ are strictly increasing continuously differentiable functions with infimum 0 and supremum 1. Thus the equation $V(a, \mathbf{b}) = k$ has a unique solution for $0 \le k \le q$ if all b_j are finite.

To illustrate equating, consider the following equating procedures for a conversion from a total score S_i to a total score $S_i' = \sum_{j=1}^{q'} X_{ij}'$. With joint maximum likelihood, for each total score $S_i = k$, $0 \le k \le q$, let $\hat{\theta}_{kS}$ satisfy

$$V(\hat{\theta}_{kS}, \hat{\boldsymbol{\beta}}_C) = k,$$

so that $\hat{\theta}_{kS}$ is the joint maximum-likelihood estimate of θ_i for any examinee i for whom $S_i = k$. With probability 1, $\hat{\theta}_{kS}$ converges to θ_{kS} (Haberman, 2004). Let the conversion from $S_i = k$ to S_i' be

$$\hat{W}_k = T'(\hat{\theta}_{kS} + \hat{\mu}' - \hat{\mu}, \hat{\boldsymbol{\beta}}').$$

With conditional maximum likelihood, for each total score $S_i = k$, $0 \le k \le q$, let $\hat{\theta}_{kC}$ satisfy

$$V(\hat{\theta}_{kC}, \hat{\boldsymbol{\beta}}) = k,$$

so that $\hat{\theta}_{kC}$ is an estimate of θ_i for $S_i = k$ that is based on conditional maximum likelihood. Let

$$V(\theta_{kC}, \boldsymbol{\beta}) = k.$$

With probability 1, $\hat{\theta}_{kC}$ converges to θ_{kC} (Haberman, 2004). Let the conversion from $S_i = k$ to S'_i be

$$\hat{W}_{kC} = V'(\hat{\theta}_{kC} + \hat{\mu}'_C - \hat{\mu}_C, \hat{\boldsymbol{\beta}}'_C).$$

These definitions are not without complications if some estimates of item parameters are infinite, but these problems can be ignored for the purpose of derivation of large-sample results.

Let

$$W_k = V'(\theta_{kS} + \nu' - \nu, \boldsymbol{\gamma}')$$

and

$$W_{kC} = V'(\theta_{kC}, \boldsymbol{\beta}').$$

It is easily shown that the probability is 1 that $\hat{W}_k$ converges to W_k and $\hat{W}_{kC}$ converges to W_{kC}. Here W_{kC} is the conversion that results if all item parameters are unknown, so it may be regarded as the correct conversion. The issue is the extent that W_k and W_{kC} differ. Only the case $1 \le k \le q-1$ is of interest, for $W_0 = W_{0C} = 0$ and $W_q = W_{qC} = q'$.

If $|\boldsymbol{\beta}|$ and $|\boldsymbol{\beta}'|$ are small, then Theorem 1 and the implicit function theorem may be applied. Let $\bar{\beta}$ be the arithmetic mean of β_j for $1 \le j \le q$, and let $\bar{\beta}'$ be the arithmetic mean of β'_j for $1 \le j \le q'$. The following results are obtained. For any real $\delta > 0$, real $\epsilon > 0$ can be found so that if $|\boldsymbol{\beta}| < \epsilon$ and $|\boldsymbol{\beta}'| < \epsilon$, then the following relationships hold:

$$|\theta_{kS} - \log[k/(q-k)] - \bar{\beta}| < \delta|\boldsymbol{\beta}|,$$

$$|\theta_{kC} - \log[k/(q-k)] - [q/(q-1)]\bar{\beta}| < \delta|\boldsymbol{\beta}|,$$

$$|\nu - [q/(q-1)]\mu| < \delta|\boldsymbol{\beta}|,$$

$$|\nu' - [q'/(q'-1)]\mu| < \delta|\boldsymbol{\beta}'|,$$

$$|W_{kC} - kq'/q - q'(k/q)(1-k/q)(\bar{\beta} - \bar{\beta}')| < \delta(|\boldsymbol{\beta}| + |\boldsymbol{\beta}'|),$$

and

$$|W_k - W_{kC} - B| < \delta(|\boldsymbol{\beta}| + |\boldsymbol{\beta}'|),$$

where

$$B = q'(k/q)(1 - k/q)[(\bar{\beta} - \mu)/(q - 1) - (\bar{\beta}' - \mu)/(q' - 1)]$$

is the approximate bias in the conversion from use of joint maximum likelihood. For the case of forms of equal length ($q = q'$), $B = (k/q)(1-k/q)[q/(q-1)](\bar{\beta} - \bar{\beta}')$ thus depends on the difference $\bar{\beta} - \bar{\beta}'$ in the average item parameters for the two forms. The suggestion is that bias is most likely to be a problem when $\bar{\beta}$ and $\bar{\beta}'$ are somewhat different and when k is about $q/2$.

To illustrate results, consider $q = q' = 20$, $k = 10$, and $\bar{\beta} - \bar{\beta}' = 1$. Then the bias approximation B is 0.2625. For the specific case of $\beta_j = \beta'_j = 0$ for $1 \leq j \leq r = 10$, $\beta_j = 1$ for $11 \leq j \leq 20$, $\beta'_j = -1$ for $11 \leq j \leq 20$, and θ_i and θ'_i both uniformly distributed on the 10 points $(h - 5.5)/10$ for $1 \leq h \leq 10$, θ_{kS} is 0.5264, θ_{kC} is 0.5, W_k is 14.58, W_{kC} is 14.40, and the actual asymptotic bias for the conversion at k is 0.1771. If β_j is changed to 0.1 for $j > 10$ and β'_j is changed to -0.1 for $j > 10$, then the new approximation $B = 0.02625$ is somewhat better, for the correct asymptotic bias is then 0.02618. The bias approximation B can be rather inaccurate for larger values of $|\boldsymbol{\beta}|$ and $|\boldsymbol{\beta}'|$. For example, for $q = 20$, $k = r = 10$, $\beta_j = \beta'_j = 0$ for $1 \leq j \leq 10$, $\beta_j = 2$ and $\beta'_j = -2$ for $11 \leq j \leq 20$, and $\theta_i = \theta'_i = 0$ with probability 1, B is 0.523, but $\theta_{kS} = 1.057$, $\theta_{kC} = 1$, $W_k = 17.02$, $W_k = 16.84$, and the asymptotic bias for the conversion is 0.182. On the other hand, a change of k to 3 leads to B equals 0.268, but $\theta_{kS} = -1.049$, $\theta_{kC} = -1.069$, $W_k = 10.03$, and $W_{kC} = 9.73$. The asymptotic bias is then 0.302.

Rather extreme examples can be constructed. Consider $q = q' = 20$, $r = 10$, $\beta_j = \beta'_j = 0$ for $1 \leq j \leq 10$, and $\beta_j = \beta'_j = 10$ for $11 \leq j \leq 20$. Thus the forms are identical in behavior. Let $\theta_i = 5$ with probability 1, and let $\theta'_i = 10$ with probability 1. Consider $k = 15$. Then $\theta_{kC} = 14.882$, $W_{kC} = 15$, and $W_k = 19.90$. Thus the conversion based on joint maximum likelihood is remarkably distant from the correct conversion for this value of k. The result is especially striking given that the forms do not differ in terms of item parameters.

At the other extreme, it is obviously true that no asymptotic bias exists if θ_i and θ'_i have the same distribution and $\beta_j = \beta'_j$ for $1 \leq j \leq q$.

To interpret these results requires some consideration of what is an unacceptable size for a bias. One criterion involves mean-squared error. By this standard, any asymptotic bias is eventually unacceptable when estimation without asymptotic bias is available. This standard is relevant to use of joint estimation, for $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})$ and $(n')^{1/2}(\hat{\boldsymbol{\beta}}' - \boldsymbol{\gamma}')$, $n^{1/2}(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta})$,

$(n')^{1/2}((\hat{\boldsymbol{\beta}}'_C - \boldsymbol{\beta}'_C)$, $n^{1/2}(\hat{\theta}_{kS} - \theta_{kS})$, and $n^{1/2}(\hat{\theta}_{kC} - \theta_{kC})$, $1 \leq k \leq q-1$ all have approximate normal distributions for n and n' large (Haberman, 1977, 2004). By standard large-sample theorem, constants b_k, b'_k, b_{kC}, and b'_{kC} exist such that $(\hat{W}_k - W_k)/(b_k/n + b'_k/n')^{1/2}$ and $(\hat{W}_{lC} - W_{kC})/(b_{kC}/n + b'_{kC}/n')^{1/2}$ have approximate standard normal distributions for n and n' large.

For one illustration, let $\sigma^2(a,b) = P(a,b)[1 - P(a,b)]$ for extended real a and b. For q large and $j > 1$, $n^{1/2}(\hat{\beta}_j - \beta_j)$ has asymptotic variance of approximately

$$1/E(\sigma^2(\theta_i, 0)) + 1/E(\sigma^2(\theta_i, \beta_j).$$

For $\beta_j = 1$, $n = 100,000$, and $\theta_i = 0$ with probability 1, the asymptotic standard deviation of $\hat{\beta}_j$ is about

$$\{[1/\sigma^2(0,0) + 1/\sigma^2(0,1)]/100000\}^{1/2} = 0.0095,$$

a value somewhat smaller than the typical asymptotic bias for $\hat{\beta}_j$.

5 Conclusion

The logical consequence of results in this report is that joint estimation is very difficult to justify for assessments of customary length. This conclusion can be awkward given that clients used to joint estimation may be reluctant to make changes; however, the biases involved are not negligible albeit not usually very large. In this report, no studied equating conversion had an asymptotic bias greater than about 0.3 except for the extreme case with $\beta_j = \beta_j = 10$ for $11 \leq j \leq 20$. The mapping of raw scores to scale scores on the reference form obviously affects the equating implications for reported scores, so implications for reported scales are not entirely predictable. Some change in reported scores can sometimes occur even for relatively small changes in conversions due to the effects of rounding, and such changes, if they occur, can normally be expected to all be in the same direction. If each raw score corresponds to a different scaled score and if simple rounding is used with the raw to raw conversions to establish a scale score, then a bias of about 0.3 is likely to result in a change in the reported scale for a W_{kC} that is from 0.2 to 0.5 greater than the nearest integer. The likelihood of a change is much lower in other cases.

It should be noted that repeated use of equating may cause bias to be a more significant issue than with one equating, especially in cases such as vertical linking in which item parameters may vary considerably across all forms to be linked. A second obvious issue is that statistical inference procedures such as approximate confidence intervals and tests of goodness of fit do not apply when asymptotic biases are present. Thus use of joint estimation prevents proper study of such basic issues as whether the Rasch model fits the data at all and how precisely item difficulties are known. Whether conditional or joint estimation is used, true-score equating with the Rasch model can not be readily justified if the Rasch model fits the data poorly.

The most fundamental issue is that no statistical or computational reason exists at the present time not to use conditional maximum likelihood and to avoid entirely the problem of asymptotic bias.

References

Andersen, E. B. (1973). *Conditional inference and models for measuring.* Copenhagen: Mentalhygiejnisk Forskningsinstitut.

Fischer, G. H. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model. *Psychometrika*, *46*, 59–77.

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, *5*, 815–841.

Haberman, S. J. (2004). *Maximum likelihood for the Rasch model for binary responses* (Research Rep. No. RR-04-20). Princeton, NJ: ETS.

Jansen, P. G. W., Wollenberg, A. L. van den, & Wierda, F. W. (1988). Correcting unconditional parameter estimates in the Rasch model for inconsistency. *Applied Psychological Measurement*, *12*, 297–306.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Wollenberg, A. L. van den, Wierda, F. W., & Jansen, P. G. W. (1988). Consistency of Rasch model parameter estimation: A simulation study. *Applied Psychological Measurement*, *12*, 307–313.

Wright, B. D. (1988). The efficacy of unconditional maximum likelihood bias correction: Comment on Jansen, van den Wollenberg, and Wierda. *Applied Psychological Measurement*, *12*, 315–318.

Wright, B. D., & Douglas, G. D. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement*, *37*, 573–586.

Appendix

Proof of Theorem 1

Let $\Gamma(k)$, $0 \le k \le q$, be the set of q-dimensional vectors $\mathbf{x}$ with coordinates 0 or 1 such that the sum of the coordinates is k, so that $\Gamma(k)$ has $q!/[k!(q-k)!]$ elements, and the conditional probability that $S_i = k$ given that $\theta_i = a$ is

$$p_{k|a}^{S|\theta} = \sum_{\mathbf{x}\in\Gamma(k)} \prod_{j=1}^{q} \frac{\exp(x_i(a-\beta_j))}{1+\exp(a-b_j)}.$$

Let $\tau_j(k)$ be $P(\theta_{kS}, \gamma_j)$ for $1 \le j \le q$ and $0 \le k \le q$. Then the conditional expected value of $\tau_j(S_i)$ given $S_i = k$ and $\theta_i = a$ is

$$Q_j(a) = \sum_{k=0}^{q} P(\theta_{kS}, \gamma_j) p_{k|a}^{S|\theta},$$

and (1) reduces to the equation

$$E(Q_j(\theta_i)) = E(P(\theta_i, \beta_j)),\ 1 \le j \le q. \tag{A1}$$

If $\boldsymbol{\beta} = \mathbf{0}$, the q-dimensional vector with coordinates 0, then the conditional distribution of S_i given $\theta_i = a$ is a binomial distribution with sample size q and probability $P(a, 0)$, and p_j is $E(P(\theta_i, 0))$ for $1 \le j \le q$. Equations (2 and A1) hold for $\boldsymbol{\gamma} = \mathbf{0}$, $\theta_{S0} = -\infty$, $\theta_{Sq} = \infty$, and $\theta_{kS} = \log[k/(q-k)]$, $1 \le k \le q-1$, for $\tau_j(k)$ is then k/q for $1 \le j \le q$ and $0 \le k \le q$, and $Q_j(a)$ is $P(a, 0)$ for $1 \le j \le q$.

The vector $\boldsymbol{\gamma}$ is a continuously differentiable function of $\boldsymbol{\beta}$ (Haberman, 2004). Let γ_{jh}, $1 \le j \le q$, be the partial derivative of γ_j with respect to β_h, and let θ_{kSh} be the partial derivative of θ_{kS}, $1 \le k \le q-1$, with respect to β_h for $2 \le h \le q$. Obviously the constraint $\gamma_1 = 0$ implies that $\gamma_{1h} = 0$. As in section 4, let $\sigma^2(a, b) = P(a, b)[1 - P(a, b)]$ for extended real a and b. Let δ_{ab} be 1 for $a = b$ and 0 otherwise. Let Z_{ijk}, $1 \le i \le n$, $1 \le j \le q$, $0 \le k \le q$, be 1 for $X_{ij} = 1$ and $S_i = k$, and let Z_{ijk} be 0 otherwise. Let Y_{ik} be 1 for $S_i = k$ and 0 otherwise. Differentiation of (A1) and (2) shows that, for $2 \le h \le q$,

$$\begin{aligned}
&\sum_{k=1}^{q-1} \sigma^2(\theta_{kS}, \gamma_j)(\theta_{kSh} - \gamma_{jh}) p_k^S \\
&\quad + \sum_{k=1}^{q} P(\theta_{kS}, \gamma_j) E(Z_{ihk} - Y_{ik} P(\theta, \beta_h)) \\
&\quad + E(\sigma^2(\theta_i, \beta_j))\delta_{jh} \\
&\quad = 0
\end{aligned}$$

for $1 \le j \le q$ and

$$\sum_{j=1}^{q} \sigma^2(\theta_{kS}, \gamma_j)(\theta_{kSh} - \gamma_{jh}) = 0$$

for $1 \le k \le q - 1$.

In the special case of $\boldsymbol{\beta} = \mathbf{0}$, θ_{kSh} is the average $q^{-1} \sum_{j=1}^{q} \gamma_{jh}$ of the γ_{jh} for $1 \le k \le q$. For each integer k from 0 to q, the distribution of $\mathbf{X}_i$ given $S_i = k$ is symmetric. Thus $E(Z_{ihk})$ is kP_k^S/q. Given that $P(\theta_{kS}, \gamma_j) = k/q$, after division by $E(\sigma^2(\theta_i, 0))$, standard results related to the binomial distribution yield the equation

$$\frac{q-1}{q}\left(q^{-1}\sum_{j'=1}^{q} \gamma_{j'h} - \gamma_{jh}\right) + \delta_{jh} - q^{-1} = 0$$

for $2 \le h \le q$. Given the constraint that $\gamma_{1h} = 0$ for $2 \le h \le q$, it follows that $q^{-1} \sum_{j=1}^{q} \gamma_{jh} = 1/(q-1)$ and $\gamma_{jh} = [q/(q-1)]\delta_{jh}$. The theorem follows from the definition of differentiability.