UNIVERSITÀ DEGLI STUDI DI UDINE
DIPARTIMENTO DI SCIENZE ECONOMICHE E STATISTICHE

IRT Methods for Chain and Multiple Equating

Michela Battauz

November 2011

n. 1/2011

IRT Methods for Chain and Multiple Equating

Michela Battauz

November 17, 2011

Abstract

Linkage plans could be rather complex, including many forms, several links and the connection of forms through different paths. This article studies item response theory equating methods for complex linkage plans when the common-item nonequivalent group design is used. An efficient way to average equating coefficients that link the same two forms through different paths will be presented and the asymptotic standard error of indirect and average equating coefficients are derived. The methodology is illustrated using simulations studies and a real data example.

Keywords: asymptotic standard errors, double equating, equating coefficients, item response theory, multiple equating, weighted bisector.

1 Introduction

When several forms need to be equated, test equating can be performed by using different linkage plans (Kolen & Brennan, 2004, §8.2.2) The linkage plan regards the choice of the forms that present a direct link and it should be developed considering various practical issues that affect the quality of the equating process (Kolen & Brennan, 2004, p. 283). For example, long chains should be avoided as they increase the amount of equating error. On the other hand, the number of links to the same form should be contained in order to preserve test security by limiting the exposure of the items. Furthermore, two old forms could be used to equate new forms in order to achieve greater equating stability and to reduce the equating error. This process is referred to as double linking. When the common-item equating to a calibrated pool design (Kolen & Brennan, 2004, §6.9.1) is used, a new form is equated to a pool of items that come from several old forms. This process originates then the linkage of the new form to two or more old forms (multiple linking). Finally, there are situations is which links are not carefully planned before test administration and the equating is performed later. For all these reasons, the linkage plan could become quite complex, involving many links, chains and alternative conversions through different forms.

The purpose of this article is to study item response theory (IRT) equating methods for complex linkage plans when the common-item nonequivalent group design is used. Both methods based on moments of item parameters, such as the mean-sigma, the mean-mean and the mean-geometric mean methods, and response function (also referred to as "characteristic curve") methods will be considered. IRT equating coefficients that link two forms through a chain of forms will be presented. Furthermore, when two forms are connected through several paths the equating coefficients related to these paths could be averaged to obtain synthetic coefficients (Kolen & Brennan, 2004, p. 280). Holland & Strawderman (2011) discussed the averaging of equating functions and showed that the angle bisector method satisfies some desirable properties. Furthermore, they generalized this method in order to include weights. A contribution of this paper is to provide an efficient way to determine the weights for averaging the equating coefficients.

The accuracy of the equating process is typically assessed by using the asymptotic standard errors of the estimators of the linking coefficients (Ogasawara, 2011). Ogasawara (2000) provided the asymptotic standard errors of IRT equating coefficients using moments, Ogasawara (2001b) gave the asymptotic standard errors of IRT equating coefficients using response function methods, while Ogasawara (2001a) derived the asymptotic standard errors of IRT true score equating and Ogasawara (2003) obtained the asymptotic standard errors of IRT observed score equating. However, all these articles consider the case of two forms to be equated. The present paper derives the asymptotic standard errors of IRT equating coefficients when several forms are equated and the linkage plan involves chains and double or multiple linking. More specifically, this paper derives asymptotic standard errors of IRT equating coefficients resulting from a chain of forms and from averaging coefficients related to different paths.

The paper is structured as follows. Section 2 illustrates IRT test equating in case of direct links and equating chains, Section 3 concerns averaging equating coefficients that link two forms through different paths and propose an efficient method to determine weights. The performance of the methodology is assessed through a simulation study in Section 4 and a real data example in Section 5. Finally, a discussion is given in Section 6.

2 IRT test equating

Consider a single test form that is denoted by g. In the three-parameter logistic model (Linden & Hambleton, 1997), the probability of a positive response on item

j in form g for a person with ability θ is given by

$$p_{gj}(\theta_{(g)}; a_{gj}, b_{gj}, c_{gj}) = c_{gj} + (1 - c_{gj}) \frac{\exp\left[Da_{gj}(\theta_{(g)} - b_{gj})\right]}{1 + \exp\left[Da_{gj}(\theta_{(g)} - b_{gj})\right]}, \tag{1}$$

where a_{gj} is the item discrimination parameter, b_{gj} is the item difficulty parameter, c_{gj} is the item guessing parameter and D is a constant typically set to 1.7. We define the parameter vector of form g as $\boldsymbol{\alpha}_g = (\boldsymbol{\alpha}_{g1}^\top, \ldots, \boldsymbol{\alpha}_{gn_g}^\top)^\top$, where $\boldsymbol{\alpha}_{gj} = (a_{gj}, b_{gj}, c_{gj})^\top, j = 1, \ldots, n_g$, and n_g is the number of items of form g. Item parameters are estimated separately for each form by using the marginal maximum likelihood method (Bock & Aitkin, 1981), regarding the person parameter θ as a random variable with standard normal distribution.

2.1 Direct equating

Let $g-1$ be another form that presents $n_{g-1\,g}$ items in common with form g. The parameters estimated for form $g-1$ can be transformed to the scale of form g by using the following equations

$$\theta_g = A_{g-1\,g}\theta_{g-1} + B_{g-1\,g}\,, \tag{2}$$

$$a_g = \frac{a_{g-1}}{A_{g-1\,g}}\,, \tag{3}$$

and

$$b_g = A_{g-1\,g}b_{g-1} + B_{g-1\,g}\,, \tag{4}$$

where $A_{g-1\,g}$ and $B_{g-1\,g}$ are the equating coefficients. These coefficients can be estimated by using moments of item parameters (Kolen & Brennan, 2004, §6.3.2; Ogasawara, 2011), or response function methods (Kolen & Brennan, 2004, §6.3.3; Ogasawara, 2001b).

Using the delta method, Ogasawara (2000) and Ogasawara (2001b) derived the asymptotic variance-covariance matrix for the vector $(A_{g-1\,g}, B_{g-1\,g})^\top$, that is given by

$$\mathrm{acov}(A_{g-1\,g}, B_{g-1\,g})^\top = \frac{\partial (A_{g-1\,g}, B_{g-1\,g})^\top}{\partial \boldsymbol{\alpha}_{g-1\,g}^\top} \mathrm{acov}(\boldsymbol{\alpha}_{g-1\,g}) \frac{\partial (A_{g-1\,g}, B_{g-1\,g})}{\partial \boldsymbol{\alpha}_{g-1\,g}}$$

where $\boldsymbol{\alpha}_{g-1\,g} = (\boldsymbol{\alpha}_g^\top, \boldsymbol{\alpha}_{g-1}^\top)^\top$ is a vector containing all the item parameters related to forms $g-1$ and g and $\mathrm{acov}(\boldsymbol{\alpha}_{g-1\,g})$ is the asymptotic variance-covariance matrix of $\boldsymbol{\alpha}_{g-1\,g}$. The derivatives depend on the method used to determine the equating coefficients and are given Ogasawara (2000) and in Ogasawara (2011) for methods based on moments, and in Ogasawara (2001b) for response function methods.

2.2 Equating chains

Suppose that two forms are linked through a chain of tests that present common items in pairs. Define the path from form 0 to form l as $p = \{0, 1, \ldots, l\}$. Applying equation (2) recursively, it is possible to obtain the the equating coefficients transforming the scale of θ_0 to that of θ_l, that are

$$A_p = A_{0,1,\ldots,l} = \prod_{g=1}^{l} A_{g-1\,g}$$

and

$$B_p = B_{0,1,\ldots,l} = \sum_{g=1}^{l} B_{g-1\,g}\; A_{g,\ldots,l}$$

where $A_{g,\ldots,l} = \prod_{h=g+1}^{l} A_{h-1\,h}$ is the coefficient that links form g to form l. These coefficients will be referred to as indirect equating coefficients.

Similarly to the case of a direct link, the delta method can be exploited to obtain the asymptotic variance-covariance matrix of the vector $(A_p, B_p)^\top$ that is

$$\mathrm{acov}(A_p, B_p)^\top = \frac{\partial (A_p, B_p)^\top}{\partial \boldsymbol{\alpha}_p^\top} \mathrm{acov}(\boldsymbol{\alpha}_p) \frac{\partial (A_p, B_p)}{\partial \boldsymbol{\alpha}_p}, \tag{5}$$

where $\boldsymbol{\alpha}_p = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top, \ldots, \boldsymbol{\alpha}_l^\top)^\top$ is the vector containing all the item parameters related to the forms that compose the path. The derivatives are given in Appendix A.

3 Average equating coefficients

Suppose that two forms are linked throw different paths. In case of two paths that link two forms the process is called double linking (Kolen & Brennan, 2004, p. 279). Define the set of paths that link two forms 0 and l as $\mathcal{P}_{0l}$ and the linking coefficients related to path p as A_p and B_p, $p \in \mathcal{P}_{0l}$. As observed by Kolen & Brennan (2004, p. 280) and Braun & Holland (1982, p. 44), the equating relationships provided by each path could be averaged to produce a single conversion that is expected to be more accurate. Holland & Strawderman (2011) discussed how to average the equating functions obtained by using different equating methods and listed seven desirable properties. They pointed out that the (weighted) mean of the equating functions satisfies all properties except symmetry, that is satisfied only in special circumstances. The symmetry property requires that the inverse function of the average equating function equals the average of the inverse functions. This property is instead satisfied by the angle bisector (Holland & Strawderman, 2011) that, in case of two linear equating functions that intersect at a point, is the linear function that bisects the angle between them.

The proposal here is to use the theory developed by Holland & Strawderman (2011) for averaging the equating functions that derive from different paths. Suppose that there are two paths that link form 0 to form l and the equations that transform the scale of θ_0 to that of θ_l are

$$\theta_l^p = A_p\, \theta_0 + B_p$$

and

$$\theta_l^b = A_b\, \theta_0 + B_b,$$

with $p, b \in \mathcal{P}_{0l}$. The angle bisector is the weighted average given by

$$\theta_l^* = w\theta_l^p + (1-w)\theta_l^b,$$

where

$$w = \frac{(1+A_p^2)^{-1/2}}{(1+A_p^2)^{-1/2} + (1+A_b^2)^{-1/2}}.$$

According to Holland & Strawderman (2011), the angle bisector can be generalized to include weights n_p, $p \in \mathcal{P}_{0l}$, and to consider more than two equations

$$\theta_l^* = \sum_{p \in \mathcal{P}_{0l}} w_p\, \theta_l^p, \tag{6}$$

where

$$w_p = \frac{n_p(1+A_p^2)^{-1/2}}{\sum_{b \in \mathcal{P}_{0l}} n_b(1+A_b^2)^{-1/2}}.$$

The average equating coefficients are then

$$A_{0l}^* = \sum_{p \in \mathcal{P}_{0l}} A_p w_p$$

and

$$B_{0l}^* = \sum_{p \in \mathcal{P}_{0l}} B_p w_p.$$

A contribution of the present paper is to derive the asymptotic variance-covariance matrix of the average equating coefficients and to develop a method to determine the weights n_p.

Note that A_p and A_b, $p, b \in \mathcal{P}_{0l}$, may be correlated because parts of the two paths may be common. The asymptotic variance-covariance matrix of the vector $(A_{0l}^*, B_{0l}^*)^\top$ can then be again obtained by using the delta method, that is

$$\text{acov}(A_{0l}^*, B_{0l}^*)^\top = \frac{\partial (A_{0l}^*, B_{0l}^*)^\top}{\partial \boldsymbol{\alpha}^\top} \text{acov}(\boldsymbol{\alpha}) \frac{\partial (A_{0l}^*, B_{0l}^*)}{\partial \boldsymbol{\alpha}},$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_p)_{p\in\mathcal{P}_{0l}}$ is the vector containing all the item parameters used in the equating process in at least one of the paths in $\mathcal{P}_{0l}$. As the weights w_p are function of A_p, the derivatives can be obtained by using the chain rule

$$\frac{\partial(A_{0l}^*, B_{0l}^*)^\top}{\partial\boldsymbol{\alpha}^\top} = \frac{\partial(A_{0l}^*, B_{0l}^*)^\top}{\partial(\mathbf{A}^\top, \mathbf{B}^\top)} \frac{\partial(\mathbf{A}^\top, \mathbf{B}^\top)}{\partial\boldsymbol{\alpha}^\top}, \tag{7}$$

where $\mathbf{A} = (A_p)_{p\in\mathcal{P}_{0l}}$ and $\mathbf{B} = (B_p)_{p\in\mathcal{P}_{0l}}$ are the vectors containing all the equating coefficients of the paths that link forms 0 and l.

When the weighted average is used, $w_p = n_p$ in Equation (6), and the derivatives are more simply given by

$$\frac{\partial(A_{0l}^*, B_{0l}^*)^\top}{\partial\boldsymbol{\alpha}^\top} = \sum_{p\in\mathcal{P}_{0l}} \frac{\partial(A_p, B_p)^\top}{\partial\boldsymbol{\alpha}^\top} n_p.$$

The derivatives are given in Appendix A.

A further issue concerns which weights n_p to use in averaging. Since equating coefficients related to different paths generally have different variability, it is reasonable to weight the coefficients in order to obtain an efficient average coefficient. The proposal of this article is to determine weights by minimizing an objective function that is the average variance of θ_l^*, namely

$$\mathrm{E}\left[\mathrm{Var}(A_{0l}^*\theta_0 + B_{0l}^*)\right] = \mathrm{Var}(A_{0l}^*) + \mathrm{Var}(B_{0l}^*), \tag{8}$$

assuming that θ_0 has zero mean and variance equal to one. The minimization can be performed numerically.

3.1 Common-Item Equating to a Calibrated Pool

A calibrated pool is a set of items coming from different forms whose parameters are expressed on the same scale. When a new form is constructed, some items from the calibrated item pool are included. The parameters that result from estimating this new form are transformed to the scale that was established for the pool and are then included in the pool. This process is known as common-item equating to a calibrated pool design (Kolen & Brennan, 2004, §6.9.1). As a result of this process, an intricate network of connections between each single form is produced and the strength of these connections is not well understood. The methodology proposed in this paper could help in analyzing the common-item equating to a calibrated pool design.

Suppose that a pool is formed by 3 forms (labeled 1, 2 and 3) and that they enter the pool in order 1, 2 and 3. For simplicity, suppose that the discrimination parameters are all set to 1 so that the equating coefficients A are all equal to 1. Furthermore, suppose that the method used for equating is any method based on moments of item parameters. Then, the conversion of the parameters of form 2 to the scale of form 1 is $b_2 + B_{21}$. The conversion of the parameters of form 3 on the scale of the pool (composed by form 1 and 2) is $b_3 + B_{31}^*$, where

$$B_{31}^* = \frac{n_{13}B_{31} + n_{23}(B_{21} + B_{32})}{n_{13} + n_{23}}.$$

See Appendix B for the derivation. This shows that the process produces a weighted average of the direct equating coefficients. However, if the forms enter the pool in order 1, 3 and 2, the parameters of form 3 are converted on the scale of form 1 using the equation $b_3 + B_{31}$ and the parameters of form 2 are converted on that of the pool using the transformation $b_2 + B_{21}^*$, where

$$B_{21}^* = \frac{n_{12}B_{21} + n_{32}(B_{31} + B_{23})}{n_{12} + n_{32}}.$$

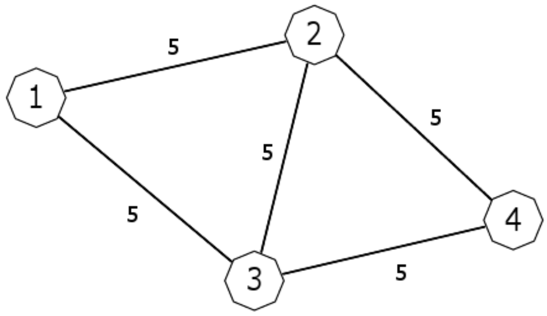
This example shows that (i) the common-item equating to a calibrated pool design produces a weighted average of the direct equating coefficients, (ii) the weights depend on the order used to form the pool, (iii) the weights are not chosen in order to obtain an efficient estimator, and (iv) the equating coefficients are not updated using the informations that derive from the subsequent forms that enter the pool. This shows that this design presents some limits and that it could be used to choose the items that form the new forms but the determination of the equating coefficients could be performed by using the methodology presented in this paper.

4 Simulation study

The performance of the chain equating coefficients, the bisector coefficients and their standard errors were assessed by means of a simulation study. In this study there are 4 forms, labeled with the numbers from 1 to 4. Each form is composed by 30 items and has 5 items in common with each of the other forms with the exception of form 1 and 4 that do not share any item. Figure 1 represents the links between the forms.

Person parameters are generated independently from a normal distribution with means varying for each form and equal to -0.3,-0.1,0.1 and 0.3 and standard deviations equal to 1.2, 1, 1 and 1.2. Item difficulty parameters were generated from a normal distribution with standard deviation equal to one. In order to

Figure 1: Linkage plan for the simulation study



obtain items with difficulties aligned with person abilities, the mean of the normal distribution was taken equal to the mean of the person parameters for the unique items, while the common items are generated from a normal distribution with mean equal to the average of the means of the person parameters. The discrimination parameters were generated from the uniform distribution with range $[0.7, 1.3]$, while the guessing parameters were set to zero. The responses where then simulated with probabilities given by equation (1) by using the `R` software (R Development Core Team, 2011). The two-parameter logistic model was fitted and the item parameters were estimated by the marginal maximum likelihood method implemented in the `ltm` package (Rizopoulos, 2006). The functions for the calculation of direct and indirect linking coefficients, weights, bisector coefficients and standard errors were programmed in the `R` statistical environment. The minimization of (8) for determining weights was performed by using the Nelder-Mead algorithm implemented in the function `optim`.

The method used for the estimation of the direct coefficients is the mean-mean method. Results are based on 100 simulations. Table 1 presents the results for the equating coefficients between form 1 and the other forms. For each pair of forms, the table reports equating coefficients obtained using direct and indirect links and the average of these coefficients obtained with the bisector and the weighted bisector methods. Weights are determined using the procedure described in Section 3. Results show that the estimators of the equating coefficients based on direct and indirect links are correct and that also averaging these coefficients by using the bisector and the weighted bisector methods provide correct estimators. Furthermore, the mean of the estimated standard errors of the coefficients are very closed to the standard deviation of the estimated coefficients. It is possible to observe that coefficients obtained from direct links present smaller standard deviation than coefficients based on indirect links and that the standard deviation increases when the chain becomes longer. The weighted bisector always provides coefficients with smaller standard deviation than the unweighted bisector. Furthermore, the standard deviation of the weighted bisector is always smaller than the standard deviation of each single coefficient used for averaging, while the unweighted bisector does not. In fact, the estimated coefficients B based on the direct link between form 1 and 2 present a standard deviation slightly smaller than the unweighted bisector coefficient.

Table 1: Results of the simulation study

link	path	A^t	B^t	mean of A	mean of B	SD of A	SD of B	mean of $se(A)$	mean of $se(B)$
12	12	1.2	-0.2	1.205	-0.206	0.079	0.071	0.088	0.081
12	132	1.2	-0.2	1.198	-0.202	0.117	0.122	0.119	0.120
12	1342	1.2	-0.2	1.207	-0.211	0.138	0.142	0.132	0.115
12	*bis*	1.2	-0.2	1.201	-0.206	0.076	0.078	0.081	0.082
12	*wbis*	1.2	-0.2	1.195	-0.204	0.067	0.066	0.072	0.072
13	123	1.2	-0.4	1.215	-0.413	0.118	0.114	0.112	0.114
13	1243	1.2	-0.4	1.210	-0.405	0.137	0.134	0.139	0.107
13	13	1.2	-0.4	1.203	-0.408	0.079	0.085	0.084	0.092
13	*bis*	1.2	-0.4	1.206	-0.408	0.076	0.075	0.081	0.078
13	*wbis*	1.2	-0.4	1.197	-0.405	0.067	0.072	0.071	0.075
14	1234	1	-0.5	1.011	-0.505	0.118	0.115	0.102	0.105
14	124	1	-0.5	1.003	-0.497	0.094	0.091	0.089	0.079
14	1324	1	-0.5	0.998	-0.495	0.118	0.123	0.117	0.110
14	134	1	-0.5	1.002	-0.501	0.093	0.096	0.093	0.087
14	*bis*	1	-0.5	1.001	-0.499	0.066	0.068	0.064	0.066
14	*wbis*	1	-0.5	0.992	-0.496	0.065	0.068	0.063	0.066

Note: *bis* denotes bisector, *wbis* denotes weighted bisector, A^t and B^t are true coefficients, mean of A and mean of B are means of estimated equating coefficients, SD of A and SD of B are standard deviations of the estimated equating coefficients, mean of $se(A)$ and mean of $se(B)$ are means of estimated standard errors.

5 Example

The methodology proposed was applied to real new data collected from a sample of students attending the third year class of high school in Italy and concern the mathematics final examination. Test developers prepared 5 forms each composed by 11 items and presenting some item in common with the others. Figure 2 represents the linkage plan and the number of items in common between the forms. Each form was administered to a different group of students composed by 418, 513, 375, 463 and 202 students. Data were analyzed using `R` software. The two-parameter IRT model was fitted and the equating coefficients using the mean-mean method and their standard error were calculated. Table 2 reports the results regarding the link between form 1 and the others. In general it is possible to observe that standard errors are quite large, indicating that the equating is not accurate. The table provides a comparison between the bisector and the weighted bisector methods for averaging with weights determined by minimizing the objective function given in Equation (8). The weighted bisector method yield smaller standard errors of the equating coefficients than the bisector method with the only exception of the standard error of the B coefficient of the link between form 1 and 2 (0.202 instead of 0.197). The weighted bisector gives always coefficients with standard errors smaller than those of each direct or indirect coefficient. Instead, the unweighted bisector in many cases provides a coefficient with standard error larger than the standard error of one of the original coefficients. For example, in the case of the link between form 1 and 4 the unweighted bisector yields coefficients with larger standard error that the direct coefficients, while the weighted bisector gives coefficient that present a gain in efficiency.

Figure 2: Linkage plan for the example.

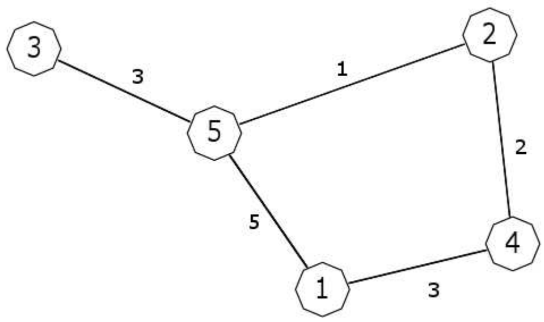


Table 2: Results of the example.

link	path	A	B	$se(A)$	$se(B)$	n_p
12	142	0.994	0.590	0.255	0.342	1.128
12	152	1.157	1.033	0.380	0.423	0.889
12	*bis*	1.072	0.802	0.226	0.197	-
12	*wbis*	1.062	0.776	0.219	0.202	-
13	14253	1.745	0.024	0.810	0.958	0.356
13	153	2.031	0.802	0.594	0.684	1.610
13	*bis*	1.880	0.390	0.605	0.645	-
13	*wbis*	1.974	0.647	0.579	0.565	-
14	14	0.649	0.117	0.100	0.128	1.600
14	1524	0.756	0.407	0.293	0.357	0.240
14	*bis*	0.701	0.259	0.150	0.156	-
14	*wbis*	0.663	0.154	0.095	0.120	-
15	1425	0.830	0.477	0.324	0.374	0.234
15	15	0.966	0.847	0.143	0.151	1.609
15	*bis*	0.896	0.656	0.183	0.177	-
15	*wbis*	0.948	0.797	0.130	0.121	-

6 Discussion

This article studies IRT equating methods for complex linkage plans and derives standard errors for chained equating coefficients and average equating coefficients. Furthermore, a method to determine weights for averaging is proposed. The simulation study and the example showed the importance of weighting in order to obtain a gain in efficiency for the average coefficients with respect the single direct and indirect coefficients, as unweighted average coefficients sometimes yield greater standard errors than the single coefficients.

This article considers standard errors of equating coefficients, however, the interest is often on standard error of adjusted scores. If a test is scored using a linear conversion of estimated IRT abilities, the calculation of standard errors of these scores is straightforward. When the true score equating method using IRT equating coefficients is chosen, the process proposed by Ogasawara (2001a) to determine asymptotic standard error of equated scores is still valid just using an appropriate acov $\left\{(\boldsymbol{\alpha}^\top, A, B)^\top\right\}$ in Equation (13) of Ogasawara (2001a). In fact, when indirect equating coefficients or average equating coefficients are used to equate two forms, the vector $\boldsymbol{\alpha}$ contains all the item parameters involved in the determination of A and B and the asymptotic variance-covariance function of $(A, B)^\top$ is that given in this article. In the same way, when the observed score equating method with equating coefficients is used, the asymptotic variance-covariance of the equated scores given in Equation (28) of Ogasawara (2003) is valid by using the asymptotic variance-covariance matrix for $(A, B)^\top$ provided in this article, where A and B can be indirect or average coefficients.

The simulation study and the example showed that sometimes the gain in efficiency provided by the weighted bisector coefficient is small with respect to one of the single coefficients used for averaging. However, averaging could be equally convenient as it can yield greater equating stability.

Appendix A: Partial Derivatives of the Equating Coefficients with respect to the Item Parameters

Indirect Coefficients

Irrespective of the method used to obtain direct equating coefficients, the partial derivatives of indirect equating coefficients used in Equation (5) are as follows.

$$\frac{\partial A_{0,\ldots,l}}{\partial a_{gj}} = A_{0,\ldots,g-1}\frac{\partial A_{g-1\,g}}{\partial a_{gj}}A_{g,\ldots,l} + A_{0,\ldots,g}\frac{\partial A_{g\,g+1}}{\partial a_{gj}}A_{g+1,\ldots,l}.$$

Note that $\frac{\partial A_{g-1\,g}}{\partial a_{gj}}$ and $\frac{\partial A_{g\,g+1}}{\partial a_{gj}}$ are both different from zero only if item j of form g is present in both form $g-1$ and in form $g+1$. Similarly,

$$\frac{\partial A_{0,\ldots,l}}{\partial b_{gj}} = A_{0,\ldots,g-1}\frac{\partial A_{g-1\,g}}{\partial b_{gj}}A_{g,\ldots,l} + A_{0,\ldots,g}\frac{\partial A_{g\,g+1}}{\partial b_{gj}}A_{g+1,\ldots,l} \ ,$$

while

$$\frac{\partial B_{0,\ldots,l}}{\partial a_{gj}} = \sum_{h=1}^{l}\left(\frac{\partial B_{h-1\,h}}{\partial a_{gj}}A_{h,\ldots,l} + B_{h-1\,h}\frac{\partial A_{h,\ldots,l}}{\partial a_{gj}}\right) \ ,$$

where $\frac{\partial B_{h-1\,h}}{\partial a_{gj}}$ is equal to zero if $g \neq h-1$ and $g \neq h$ and $\frac{\partial A_{h,\ldots,l}}{\partial a_{gj}}$ is equal to zero if $g < h$. Finally,

$$\frac{\partial B_{0,\ldots,l}}{\partial b_{gj}} = \sum_{h=1}^{l}\left(\frac{\partial B_{h-1\,h}}{\partial b_{gj}}A_{h,\ldots,l} + B_{h-1\,h}\frac{\partial A_{h,\ldots,l}}{\partial b_{gj}}\right).$$

Bisector Equating Coefficients

The partial derivatives of the bisector coefficients with respect to direct and indirect equating coefficients relative to one of the paths that link to forms used in Equation (7) are as follows:

$$\frac{\partial A_{0l}^*}{\partial A_p} = [1 - A_p^2(1 + A_p^2)^{-1}]w_p + A_{0l}^* w_p A_p (1 + A_p^2)^{-1},$$

$$\frac{\partial A_{0l}^*}{\partial B_p} = 0,$$

$$\frac{\partial B_{0l}^*}{\partial A_p} = -A_p B_p (1 + A_p^2)^{-1} w_p + B_{0l}^* w_p A_p (1 + A_p^2)^{-1},$$

$$\frac{\partial B_{0l}^*}{\partial B_p} = w_p.$$

Appendix B: Equating Coefficients for Common-Item Equating to a Calibrated Pool

It is assumed that $a_{gj} = 1$ for all g and j. Then, the equating coefficient for converting the parameters of form 2 on the scale of form 1 is given by

$$B_{12} = \frac{1}{n_{12}} \sum_{j \in I_{12}} b_{1j} - \frac{1}{n_{12}} \sum_{j \in I_{12}} b_{2j}.$$

We denote by $I_{p3} = I_{13} \cup I_{23}$ the set of items in common between the pool and form 3, and $n_{p3} = n_{13} + n_{23}$ (assuming that $I_{13} \cap I_{23} = \oslash$) the cardinality of I_{p3} and b_{pj} the j-th difficulty item parameter of the pool.

The equating coefficient for converting the parameters of form 3 on the scale of the pool is given by

$$\begin{aligned}
B_{13}^* &= \frac{1}{n_{p3}} \sum_{j \in I_{p3}} b_{pj} - \frac{1}{n_{p3}} \sum_{j \in I_{p3}} b_{3j} \\
&= \frac{1}{n_{13} + n_{23}} \left(\sum_{j \in I_{13}} b_{1j} + \sum_{j \in I_{23}} (b_{2j} + B_{12}) \right) - \frac{1}{n_{13} + n_{23}} \left(\sum_{j \in I_{13}} b_{3j} + \sum_{j \in I_{23}} b_{3j} \right) \\
&= \frac{\sum_{j \in I_{13}} b_{1j} - \sum_{j \in I_{13}} b_{3j} + \sum_{j \in I_{23}} (b_{2j} + B_{12}) - \sum_{j \in I_{23}} b_{3j}}{n_{13} + n_{23}} \\
&= \frac{n_{13} B_{13} + n_{23}(B_{12} + B_{23})}{n_{13} + n_{23}}.
\end{aligned}$$

References

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

Braun, H. I. & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 9–49). New York: Academic.

Holland, P. W. & Strawderman, W. E. (2011). How to average equating functions if you must. In von Davier A. A. (Ed.), *Statistical Models for Test Equating, Scaling, and Linking* (pp. 89–107). New York: Springer-Verlag.

Kolen, M. J. & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices (Second Edition)*. New York: Springer-Verlag.

Linden, W. J. v. d. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. Springer-Verlag.

Ogasawara, H. (2000). Asymptotic standard errors of IRT equating coefficients using moments. *Economic Review (Otaru University of Commerce)*, 51, 1–23.

Ogasawara, H. (2001a). Item response theory true score equatings and their standard errors. *Journal of Educational and Behavioral Statistics*, 26, 31–50.

Ogasawara, H. (2001b). Standard errors of item response theory equating/linking by response function methods. *Applied Psychological Measurement*, 25, 53–67.

Ogasawara, H. (2003). Asymptotic standard errors of IRT observed-score equating methods. *Psychometrika*, 68, 193–211.

Ogasawara, H. (2011). Applications of asymptotic expansion in item response theory linking. In von Davier A. A. (Ed.), *Statistical Models for Test Equating, Scaling, and Linking*. New York: Springer-Verlag.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1–25.